AD_____

Award Number: W81XWH-11-1-0525

TITLE: Identification of the Gene for Scleroderma in the Tsk/2 Mouse Strain: Implications for Human Scleroderma Pathogenesis and Subset Distinctions

PRINCIPAL INVESTIGATOR: Michael L. Whitfield, Ph.D.

CONTRACTING ORGANIZATION: Dartmouth College
Hanover, NH 03755-1404

REPORT DATE: July 2013

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2013	Annual	1 July 2012 – 30 June 2013

4. TITLE AND SUBTITLE

Identification of the Gene for Scleroderma in the Tsk/2 Mouse Strain: Implications for Human Scleroderma Pathogenesis and Subset Distinctions

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0525

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Michael L. Whitfield, Ph.D.
Elizabeth P. Blankenhorn, Ph.D.
Carol Artlett, Ph.D.
E-Mail: michael.L.whitfield@dartmouth.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Dartmouth College
Hanover, NH 03755-1404

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
We have worked together with Drs Blankenhorn and Artlett (our partnering PIs) to submit one document that provides a comprehensive summary of our joint project

14. ABSTRACT

This project focuses on an animal model of the human disease, systemic sclerosis (SSc), called Tsk2/+. The SSc-like traits in Tsk2/+ heterozygotes are highly penetrant. In addition to a readily apparent skin fibrosis resulting from ECM anomalies, Tsk2/+ mice show autoimmune and inflammatory features that closely resemble human SSc features, making it useful as a pre-clinical model. In this report, we show a clear time dependence on the gene expression in the skin of the Tsk2/+ mice. We have proven that Col3a1 is the gene that causes Tsk2/+ fibrosis, and have confirmed the sequence difference between Tsk2/+ and the parent strain, 101. We present preliminary results on the expression of TGFβ-dependent mRNA signatures from Tsk2/+ and WT littermates that suggest the mechanism for the fibrosis is related to the up-regulation of TGFβ signatures seen in the mutant strain. We show that Col3A1 fibers are altered in aged Tsk2/+ mice.

15. SUBJECT TERMS
nimal model, systemic sclerosis, scleroderma, Tsk2/+, fibrosis, gene, genetics, TGFβ

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	42	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

COVER PAGE 1

REPORT DOCUMENTATION PAGE 2

INTRODUCTION 4

BODY 4

Milestone 1 4
Milestone 2 4
Milestone 3 5
Milestone 4 6
Milestone 5 6
Milestone 6 6

PRELIMINARY RESULTS AND RESEARCH ACCOMPLISHMENTS BY MILESTONE 6

Milestone 1 6
Milestone 2 6
Milestone 3 and 5 7

KEY RESEARCH ACCOMPLISHMENTS (July 1 2012-June 30, 2013) 13

REPORTABLE OUTCOMES (July 1, 2012 to June 30, 2013) 13

Abstracts and Presentations: 13
Manuscripts: 13
Presentations: 14
Degrees obtained that are supported by this award 15
Development of cell lines, tissue or serum repositories 15

CONCLUSION: 15

REFERENCES 15

APPENDIX 15

INTRODUCTION

Scleroderma and systemic sclerosis (SSc) is a heterogeneous disease of fibrosis and inflammation, concomitant with significant autoimmunity. SSc often presents with skin manifestations and Raynaud's phenomenon; the extent and location of fibrotic lesions in people with SSc contributes to the diagnoses of disease subtypes and prognosis. Several preclinical animal models for SSc exist. Tsk2/+ mice were discovered more than two decades ago when progeny of a 101/H male in an ENU mutagenesis experiment were noted with very tight skin. *Tsk2* is homozygous lethal, similar to the *Tsk1* mouse model of SSc which results from duplication of the fibrillin gene. Tsk1 has been one of the most commonly used models for SSc and therefore has been extensively characterized. The SSc-like traits in Tsk2/+ heterozygotes are highly penetrant. In addition to a readily apparent skin fibrosis resulting from ECM anomalies, Tsk2/+ mice show more autoimmune and inflammatory features than Tsk1/+, and their longer lifespan and immune features that closely resemble human SSc features are ideal for use as a pre-clinical model. The Tsk2 mutation has been bred onto a homogeneous inbred (C57Bl/6, or B6) background in Dr. Blankenhorn's laboratory. B6.Tsk2/+ mice have many features of the human disease, including tight skin, dysregulated extracellular matrix deposition, and significant autoimmunity. We have found that Tsk2-mediated autoimmune and fibrotic signs develop *progressively* with age and manifest differently in females than males, a phenomenon also observed in human SSc. These SSc phenotypes in B6.Tsk2 mice are all likely due to a single genetic mutation, which we have now unidentified. We proposed to identify the *Tsk2* gene and understand its mechanism of action as outlined in our statement of work. This mouse affords a unique opportunity to examine the pathways leading to the multiple clinical parameters of fibrotic disease from birth onward, and we describe our studies from the second year of the grant below. This work was accomplished by researchers at Drexel University College of Medicine and Dartmouth Geisel School of Medicine under the partnering PI option.

BODY

Milestones were assigned to this proposal, with tasks to be accomplished by each investigator. The overall **summary** of our progress relative to these tasks is given below, followed by a complete discussion of our work this year.

Milestone 1 Identify Tsk2/+ gene:

Task 1 was for the Blankenhorn laboratory to collect DNA for sequencing (Months 1-6), which we have done. In year 1, we collected the Tsk2/+ and 101/H (parental strain) DNAs for sequencing on the 454.

Task 2 (Months 6-12) was to select anchor sequences for Nimblegen chip design, so that chromosome 1 DNA in the Tsk2/+ interval could be sequenced. This was done in year 1 by our subcontractors at ASRI, Dr. Fen Hu and Dr. Garth Ehrlich.

Task 3 (Month 6-12): Dr. Hu and her colleagues have hybridized the mouse genomic DNA to the chips and collected Tsk2/+ interval DNA, meeting this target. They have sequenced both Tsk2/+ and 101/H interval DNA.

Task 4: We assembled all the sequence data in years 1 and 2, and have aligned the sequences to compare and report all observed polymorphisms. At Drexel, we completed amplification and re-sequencing of the target gene this year (year 2, on schedule), Blankenhorn lab.

Milestone 2 Determination of mechanism of action of Tsk2/+ gene. Overall, this study is still in progress.

Task 1 (Months 18-32): At Drexel, we were to breed Tsk2/+ mice to a knockout mouse with a deficiency in the newly-identified *Tsk2* gene, to determine if either Tsk2 or wild-type allele can complement the genetic deficiency. Last year, we reported that *Col3A1* is the gene underlying *Tsk2,* and so the Blankenhorn laboratory purchased and bred three Col3A1 KO/WT heterozygous male mice to Tsk2/+ dams in July 2012 (month 12). We now present results from this cross in our progress for year 2, below.

Task 2 (Months 1-36). Correlate the known actions of the Tsk2 gene at Drexel with gene expression data at Dartmouth (Aim 2) and with the presence of proliferating cells (Aim 3). In this Task, largely accomplished at Dartmouth with the microarray studies (months 4-12), we are establishing the timeline for the gene signatures in male and female Tsk2/+ mice. We will then examine the corresponding Tsk2/KO mice for these phenotypes to detect alterations in the TGFβ1-driven proliferative, fibrotic signature of the Tsk2/+ gene when it is absent. This year, we have added a task to the project: we have found that mouse skin samples for gene expression studies need to be stratified for hair cycle. This has delayed some of our gene signature analyses until year 2 and 3, as the earlier samples sent to Dartmouth were not all in the same stage of the hair cycle, which dramatically alters the landscape of the skin. This is now addressed.

Also part of Milestone 2 was the aim to make fibroblast cultures from the mice. Originally planned for the appropriate conditional TGFβR animals so that alterations to TGFβ1 signaling taking place early in mouse post-natal development can be monitored, we have found this to be a good approach for the analyses of all the genotypes. A summary of progress in this year is below.

Another component of Milestone 2: Based on Dr. Whitfield's preliminary results with whole genome profiling, and on Dr. Blankenhorn's with selected TGFβ1-dependent target gene expression, we expect that TGFβ1 is a necessary component in the disease pathway. Therefore, we are ordering the TGFβR conditional KO mice and the transgenic mice bearing *cre* recombinase under the control of a collagen promoter. We should have them in hand this fall. When bred to Tsk2/+ mice, these constructs will help us to fully understand the timeline of the TGFβ signature. At Drexel, we will breed Tsk2/+ mice to TGFβR conditionally deficient mice (by breeding the Tsk2/+ mutation onto a floxed TGFβR2 KO and then breed the resulting mice to B6.Col-Cre animals) to determine the interaction between TGFβ1 and *Tsk2*. Approximately equal numbers of wild-type (WT) and Tsk2/+ mutant progeny with the dominant TGFβR conditionally deficient trait will be born and used (Months 24-36). We will then correlate the known actions of the *Tsk2* gene AT DREXEL with gene expression data AT DARTMOUTH in Aim 2 and with the presence of proliferating cells in Aim 3, by examining the corresponding Tsk2/KO mice for these phenotypes to detect alterations in the TGFβ1-driven proliferative, fibrotic signature of the Tsk2/+ gene when it is absent. Months 18-36, Blankenhorn laboratory; months 12-18 in the Whitfield laboratory

Milestone 3 Determine the timing of TGFβ activation in the Tsk2/+ mice, and differences between males and females. Dr. Blankenhorn will send mouse tissues to Dr. Whitfield, who will do the RNA work. Overall, this study is still in progress.

Task 1 (Months 1-36): The Blankenhorn laboratory will breed sufficient numbers of mice to collect skin at postnatal Day 0, day 7, day 14, day 21 as well as 1 month and 4 months. These mice are used by all three investigators, and whenever possible, each individual mouse was studied for the relevant traits in each laboratory, so that histology and RNA transcript analysis will occur on the same animal. We have met our targets in year 1. In Year 2, this task was accomplished, and in year 3, we will add samples where stratification by hair cycle has been done.

Task 2 (Months 4-12) Prepare RNA from skin at Drexel and hybridize DNA microarrays at Dartmouth. Data will be analyzed, processed and stored. In practice, we found it better to send whole skin samples to Dartmouth and prepare the RNA there. This is ongoing.

Task 3 (Months 12-36): At Dartmouth perform data analysis for expression of TGFβ as well as other gene signatures, both profibrotic (IL13 and IL4) and those that may not be expected (genome-wide). This is ongoing.

Task 4 (12-24 months, if necessary): If the microarray study is unclear, we had proposed a small number of RNASeq runs to validate the gene expression data. This was unnecessary for these goals.

Task 5 (dependent timing): Immunohistochemistry will be performed for the validation of TGFβ signatures found in the microarrays. This will be performed by the Artlett and Blankenhorn laboratories. This study relies on the completion of Aim 2 (Milestones 2 and 3), for which we need microarray data from all ages.

Milestone 4 Characterize how well the Tsk2/+ mouse approximates human SSc at different time points. These tasks are on time, and to be conducted largely in year 3.

Task 1: At Dartmouth, map mouse genes to human orthologs, integrate mouse and human data using Distance Weighted Discrimination to remove systematic biases, and cluster mouse and human data (months 12-36).

Task 2: At Dartmouth, Analyze data-driven groupings, pathways, computational validation and data interpretation (months 12-36). Data analysis for expression of proliferative signatures will give us a way to understand the subset of SSc patients that exhibit diffuse clinical symptoms with signs of cell proliferation. This is a special investigation of proliferative signatures by the Whitfield group to capitalize on their extensive experience with cell cycle and proliferative motifs in gene expression. It was scheduled for months 8-24, and is in process.

Milestone 5

Task 1: We will perform confirmation qRT-PCR on select genes based on Aim 2 in the Blankenhorn and Whitfield laboratories. We had scheduled this for months 4-24; and thus, we have not completed this task as microarray data from all ages is not yet finished due to the confounding hair cycle. In retrospect, we will modify this task to extend to year three as well, to ensure full study of interesting gene expression patterns over mouse developmental ages. In the Drexel laboratories, we plan experimentation on the mechanotension of the ECM when it contains Tsk2/+ collagen in comparison to ECM containing WT collagen, after the identification of the Tsk2/+ candidate gene by Aim 1. We have not yet started this; one issue is how slowly the mouse fibroblasts grow. A final goal within this milestone is to characterize the disorder to understand how the *Tsk2* mutation acts to elicit it.

Milestone 6 Cross-breed Tsk2/+ mice to Wsh mast cell knockout mice (at Drexel). This Milestone was *deleted* and supplanted by other work, as described in the Year 1 Progress report.

PRELIMINARY RESULTS AND RESEARCH ACCOMPLISHMENTS BY MILESTONE

Milestone 1

The nucleotide sequence of the Tsk2/+ region was accomplished in year 1 and 2, and the initial report of the sequencing capture and early resequencing at Drexel was made in the year 1 progress report. We confirmed the global sequencing result at Drexel and also evaluated the remaining SNPS by phototyping[1, 2]. Of these, only the Col3A1 non-synonymous coding SNP was validated; two intronic SNPs in the GULP1 gene also distinguish Tsk2/+ from all other strains for which chr 1 genotyping is available. The Col3a1 SNP results in a Cys to Ser change in the PIIINP (N-terminal) cleavage product of the Col3a1, and thus is a target for our research in the immediate future.

Milestone 2

Breeding Tsk2 to Col3A1 KO: We began the breeding necessary for the genetic complementation test of *Col3a1* by breeding the Tsk2/+ line to BALB.Col3A1KO mice (heterozygotes as well). Results were collected in year 2. **This study provided the definitive evidence that *Col3a1* is the gene mutated in Tsk2/+ mice**. We hypothesized that combining the Tsk2 mutation with a Col3a1- allele would not generate a viable genotype, and the Tsk2/*Col3a1*KO mice would die in utero. There are four possible genotypes in the cross: (1)Tsk2/+; (2)Tsk2/*Col3A1*KO; (3)Col3A1KO/*Col3A1*KO; and (4)*Col3A1*KO/+. Our analysis of the litters confirmed this prediction: we found no Tsk2/Col3a1- mice among the neonates (i.e., this genotype is nonviable), out of 33 offspring born in this cross (**Figure 1; Table 1**). This implies that the C33S mutation in *Col3a1* is the causative

mutation, because the Tsk2-bearing chromosome has no functional *Col3a1* gene. While the compound heterozygotes couldn't survive without a functional allele of *Col3a1* (and C33S is the only mutation within *Col3a1* in the Tsk2/+ mice), there were a few Col3A1KO/Col3A1KO homozygotes born. These mice did not thrive and were sacrificed to provide neonatal skin fibroblast cultures.

Fibroblast cultures were established in Blankenhorn and Artlett lab in months 8-24. Western blots of collagen expression were compared to hydroxyproline and Sirius red measurements. A very small but consistent and significant average difference is seen between WT and Tsk2/+ fibroblasts collected at day 1 (**Figure 2**), but by day 28 and older, this difference is difficult to repeat. Ascorbic acid supplementation did not reveal any further differences. One note – to date, we have not observed proliferation differences in Tsk2/+ vs. WT fibroblasts in vitro.

Milestone 3 and 5

RNA gene expression profiling.

We reported (in our Year 1 Progress Report) the initiation of the planned RNA gene expression profiling by DNA microarray at Dartmouth. To accomplish this milestone we have analyzed skin from both wt and Tsk2/+ mice at 4, 8, 12, and 20 weeks of age for both males and females.

We analyzed 4 independent skin samples each for WT and Tsk2/+ mice at each time point for both male (24 microarrays) and female mice (24 microarrays). Each of these was analyzed separately. We found a clear time dependence of the gene expression in Tsk2/+ that also varied by gender. Analysis of the female mice at 4, 8, 12 and 20 weeks of age identified specific gene expression signatures at each time point and it was very clear that some time points had very significant changes in gene expression (4 and 12 weeks), whereas other time points (8 and 20 weeks) show many fewer significant genes with higher False Discovery Rates (FDRs). A similar finding was observed for males except the largest changes were observed at 8 weeks and the fewest

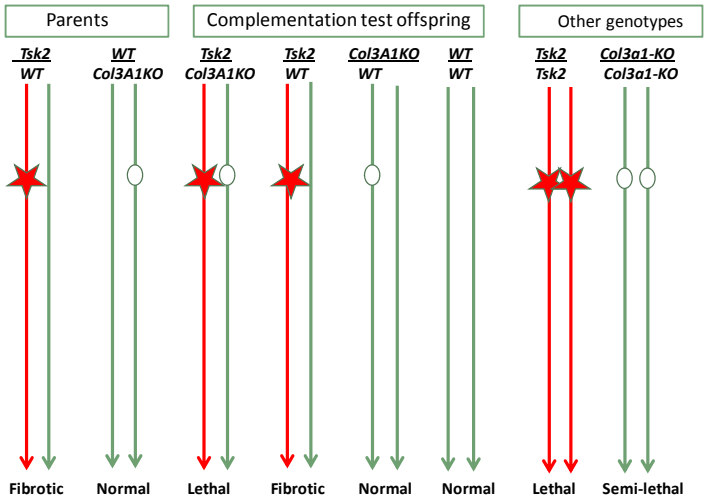


Figure 1. Complementation test proves Tsk2 is C33S: Red lines indicate the mutated chromosome from Tsk2/+ mice and green lines indicate a normal chromosome 1. Ovals:*Col3a1* knocked out in BALB/c mice; red star shows the *Tsk2* mutation. Outcomes are indicated

Parents	Tsk2/+ Tight skin	+/+ Normal skin	Tsk2/Tsk2 (lethal)
Tsk2/+ x Tsk2/+	22	21	0
	$Col3a1^{+}$/ $Col3a1^{-}$	$Col3a1^{+}$/ $Col3a1^{+}$	$Col3a1^{-}$/ $Col3a1^{-}$
$Col3a1^{-/+}$ x $Col3a1^{-}$ /+	16	13	3
	WT/$Col3a1^{+}$	Tsk2/$Col3a1^{+}$	WT/$Col3a1^{-}$	Tsk2/$Col3a1^{-}$
Tsk2/+ x $Col3a1^{-/+}$	12	10	11	0

Table 1: Offspring born of Tsk2 and Col3a1-KO mice. The breeding scheme for maintaining the two mutant lines is shown in the first two rows, with the phenotypes in the pups heading each column. The complementation test of (*Tsk2*/+ x *Col3a1*-/+) is shown in the bottom

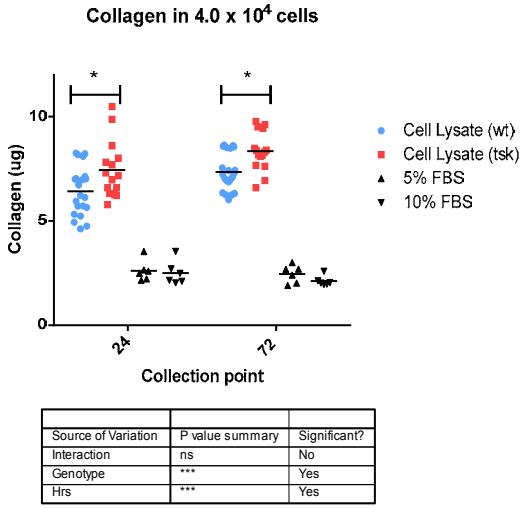


Figure 2. Collagen synthesis *in vitro*, stratified by Tsk2/+ genotype.

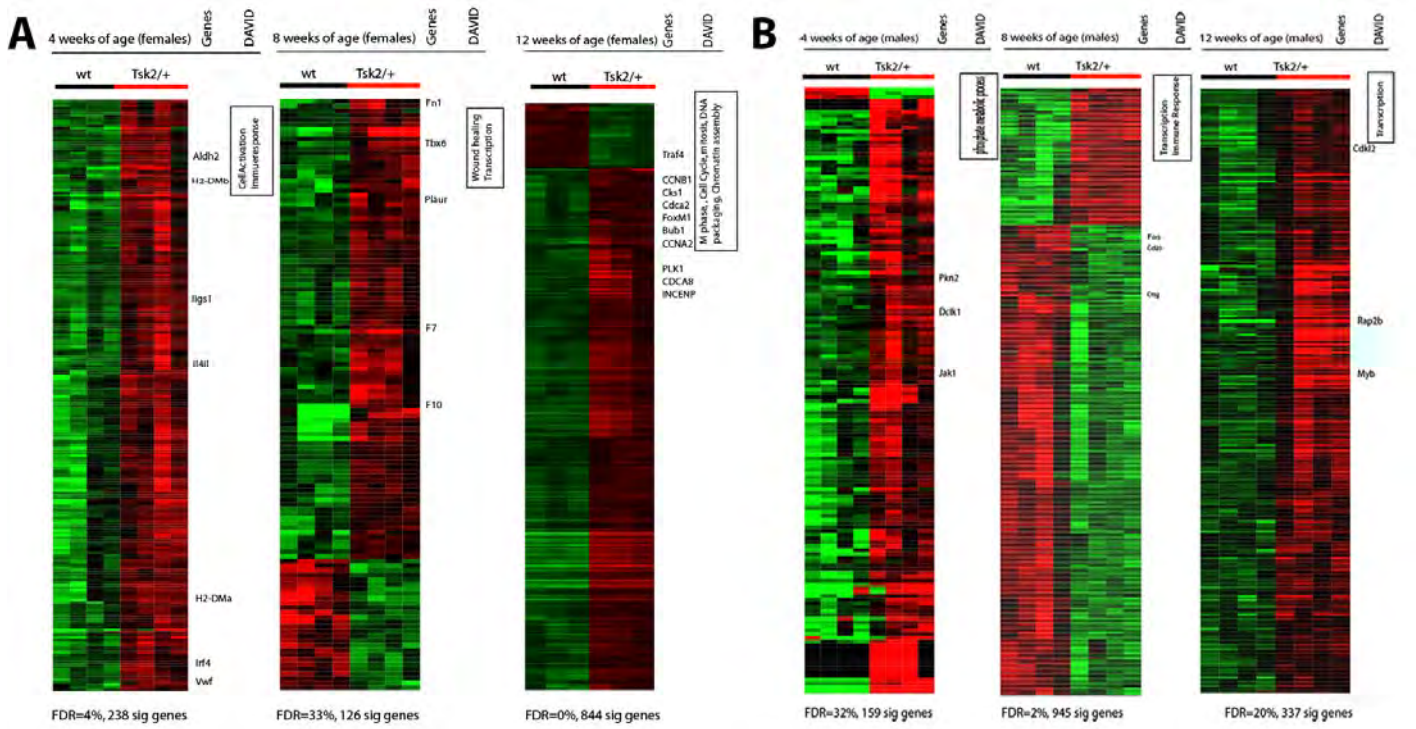


Figure 3. Genes differentially expressed between Tsk2/+ and WT differ by time point and gender. Four-wk and 12-wk female Tsk2/+ mice (panel A), as well as 8-wk male Tsk2/+ mice (panel B) show significantly more differentially genes at a low false discovery rate (FDR). We used Significance Analysis of Microarrays to compare WT and Tsk2/+ mice at different ages for each sex. **A.** 238 significant genes at 4% FDR were identified in 4-wk old females, whereas, only 126 significant genes were obtained at 8-wk (FDR=33%). However, 12-wk old females showed the most significant differential gene expression with 844 significant genes (0.1% FDR). **B.** For males the most significant difference was observed in 8 wk old animals. The 8-wk old male Tsk2/+ mice showed 945 significant genes (FDR=2). 4-wk and 12-wk male Tsk2/+ mice showed significant differential gene expression but at very high FDR. Similar results were seen with RT-PCR of selected transcripts in male and female mice of different ages.

changes observed at 4 and 12 weeks, suggesting that the gene expression changes in male Tsk2/+ mice are changing at a different rate, or at different developmental time points, than occurs in female mice (Figure 3).

Based on data from Dr. Blankenhorn's laboratory, the hypothesis was developed that hair cycle varied between Tsk2/+ and WT as well as between males and females. In addition, this process appears somewhat stochastic, occurring in different animals of the same genotype and gender at different times. We therefore have recently revisited the above analysis controlling for hair cycle stage for each sample and only using those samples that are at the same stage. We repeated the gene expression analysis and show that by eliminating those samples from mice in anagen, we increased the number of genes we are able to select, and correspondingly reduced our overall FDR due to decrease sample variation (Figure 4). The number of genes that can be selected at a given FDR is shown for each gender and age group of mice in Figure 4.

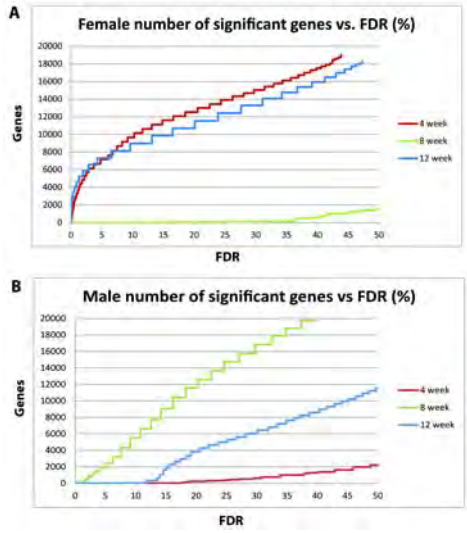


Figure 4. Number of differentially expressed genes by FDR. A. Females at each age were analyzed for genes differentially expressed between WT and Tsk2/+. The most significant changes are observed at 4 and 8 weeks. **B.** Males show the largest changes occurring at 8 and 12 weeks, and relatively few changes occurring at 4 weeks.

A Tgfβ-responsive gene signature is activated in Tsk2/+

Using samples from 4-week old female mice that show the most significant change in gene expression when we compare WT and Tsk2/+ mice (Figure 5), we then asked if there was change in

the expression of TGFβ-regulated genes in these mice. Our analysis of the Gene Ontology (GO) annotations shows that the majority of genes up-regulated in Tsk2/+ mouse skin at 4 weeks of age map to the GO Biological processes of Cell adhesion and Cell morphogenesis (DAVID, Benjamini-corrected $p < 0.05$). Genes that show increased expression include *Col6a1, Col6a2, Col5a1, Sparc* and *Thy1*. Many of these genes are known targets of the profibrotic cytokine TGFβ.

To formally identify the TGFβ responsive genes in an unbiased fashion, we used a unique 674-gene TGFβ signature that we previously developed from human dermal fibroblasts [3]. Those human genes were then mapped to 526 mouse gene orthologs using the Mouse Genome Informatics (MGI) Mouse/Human Orthology Phenotype Annotations table (The Jackson Laboratory, Bar Harbor, Maine). Out of those 526 orthologs, 459 genes were annotated on the Agilent 4X44K Mouse Whole Genome Microarray. We then removed those genes that were likely to be affected by the hair growth cycle, or background noise, and in the end used 134 genes as our final, mouse TGFβ-responsive gene signature. These 134 genes represent the core TGFβ signature and include the well-characterized TGFβ target genes.

Enrichment of the TGFβ responsive gene signature was analyzed in skin samples from 4-week old female mice that show the largest changes in gene expression (see above) and in samples that have been controlled for anagen (e.g. any sample that was in anagen was removed and only samples not in active hair cycle growth were considered). We then calculated enrichment of the TGFβ gene signature in each mouse sample by calculating the Pearson correlation coefficients between the centroid and the gene expression for each 4-week female mouse skin biopsy. These data are shown in **Figure 5** and clearly *show enrichment for TGFβ-responsive gene expression in Tsk2/+ female mice at 4 weeks of age.*

Overall, our results clearly show enrichment of the TGFβ-responsive gene signature in Tsk2/+ mice but not in WT mice, and this is dependent on age and hair cycle. We are now collecting skin samples to study even younger mice (2 weeks of age), in which we have seen a signature difference in males in preliminary analyses.

Milestone 5 Reticular fibers staining of lung tissue

We further pursued the finding of abnormal reticular fibers (type III collagen) in the tissues, and investigated lung in Tsk2/+ and wild-type littermate mice. We found that Tsk2/+ mice, in addition to having abnormal reticular fibers in the skin (2012 report), demonstrate reticular fiber abnormalities in the lung of Tsk2/+ mice at 52 weeks of age (**Figures 6-9**). We are currently investigating earlier ages to determine whether this observation holds true in younger animals. **Figure 6** shows an overview of the overall pathological differences observed between Tsk2/+ and their wild-type littermates. These are described in more detail in **Figs 7, 8, and 9**. All aspects of the lung in the Tsk2/+ mouse show an overt increase in reticular fibers.

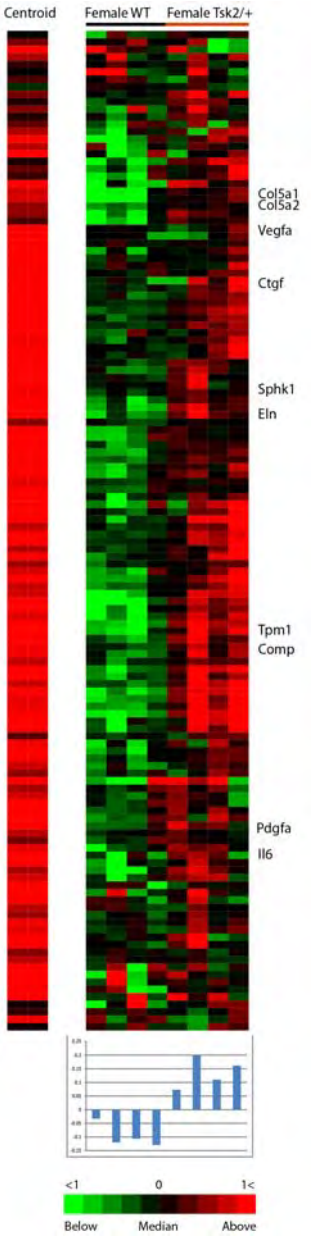


Figure 5. A TGFβ responsive signature is activated in 4-wk old female Tsk2/+ mice, not in WT. A TGFβ responsive gene signature centroid was generated from TGFβ-treated human primary fibroblast [3]. The centroid representing the average of the TGFβ responsive genes is shown to the left of the heat map. Pearson correlations between the centroid and each array were calculated and are plotted directly beneath each microarray analyzing WT or Tsk2/+ mice from 4-week old female mice. Several known Tgfβ target genes are highlighted.

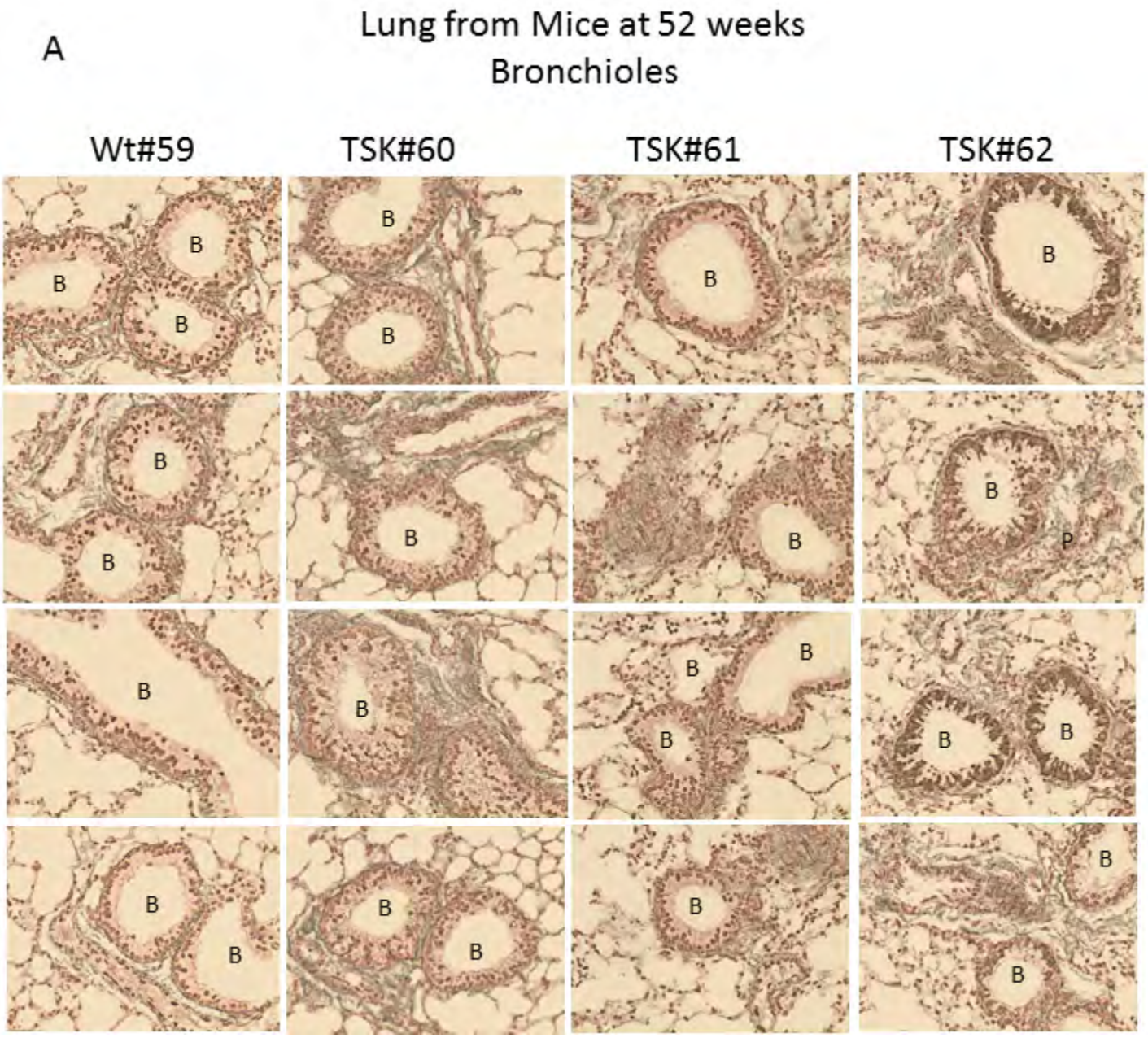


Figure 6. In the wild-type (WT#59), the reticular fibers (gray) line the bronchioles (B) tightly and are well delineated. In TSK2/+ (Tsk2#60, Tsk2#61, and Tsk2#62), the reticular fibers appear to be disorganized around the base of the bronchioles and there is more COL3A1 in the surrounding adventitia.

On closer examination of the reticular fibers surrounding the bronchioles, we observed that in the wild-type lung there was a single layer of reticular basement cells surrounded by reticular fibers that was densely stained (arrow). However in the Tsk2/+ mouse, even though the bronchioles were surrounded by a single layer of reticular basement cells (arrow), the reticular fibers were diffuse and not well defined (**Figure 7**).

We noted additional lung abnormalities in the Tsk2/+ mouse. We observed that there were increased reticular fibers surrounding the alveoli. Alveoli are normally supported by a fine network of fibers (arrows in wild-type image); however in the Tsk2/+ mouse, there was increased staining for type III collagen and we noted that this was also disorganized and not well delineated as that observed in the wild-type mouse at the same age (**Figure 8)**.

Finally we note that the smooth muscle in the lung had heavy staining in the Tsk2/+ mouse and that there were thickened areas of stain between the columnar epithelial cells (**Figure 9**). We found that the lung

parenchyma overall was normal with the exception of the increased reticular fiber staining as shown above.

In light of these findings in the lung tissue and the findings reported in the skin (2012 report), we will investigate other tissues (e.g. spleen and liver) that are enriched with reticular fibers to determine whether the fibers are also abnormal in those organs. We are currently collecting these organs at different ages from the wild-type and Tsk2/+ mice for this analysis.

Based on our finding that there is a mutation in the PIIINP region of the *Col3a1* gene, and that it has been reported that the PIIINP fragment is involved in fibrillogenesis of type I collagen, we speculated that we would not be able to extract protein from the skin of these mice as easily as from the wild-type mice using 1M NaCl or 0.5M acetic acid. Indeed, we found that less protein was extracted with 1M NaCl from the Tsk2/+ skin but the amount of extracted protein with 0.5M acetic acid was not significantly different. Acid extractable protein was found to be 0.33 mg/ml in the wild-type vs 0.32 mg/ml in Tsk2, p=0.47; whereas for salt extracted protein there was 0.86 mg/ml from the wild-type vs. 0.68 mg/ml from Tsk2, p=0.045. We are extending these findings to additional animals and want to determine why there is less extractable protein in Tsk skin. Our goal now is to establish how this mutation alters collagen fibrillogenesis leading to thickened skin and altered reticular fiber pathology.

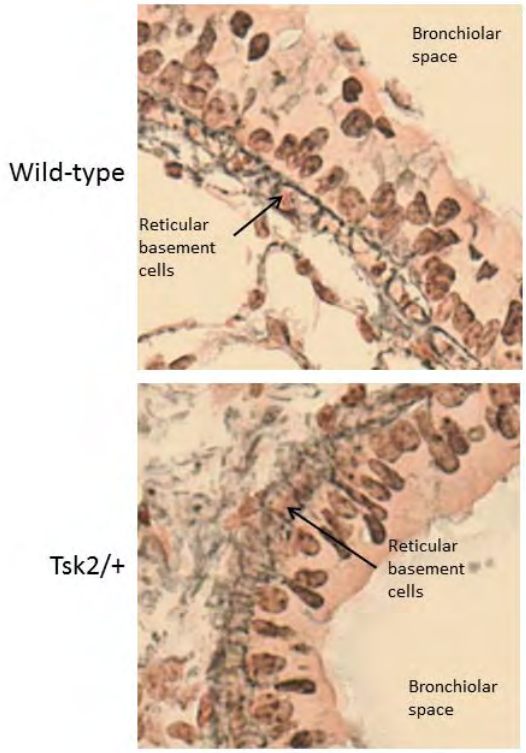


Figure 7. Reticular fiber are disorganized surrounding reticular basement cells in Tsk2/+ mice.

We are now staged in the final year to investigate altered protein synthesis of specific genes in the skin of the Tsk2/+ mouse as determined by the increased gene expression that was identified in collaboration with Dr. Michael Whitfield at Dartmouth.

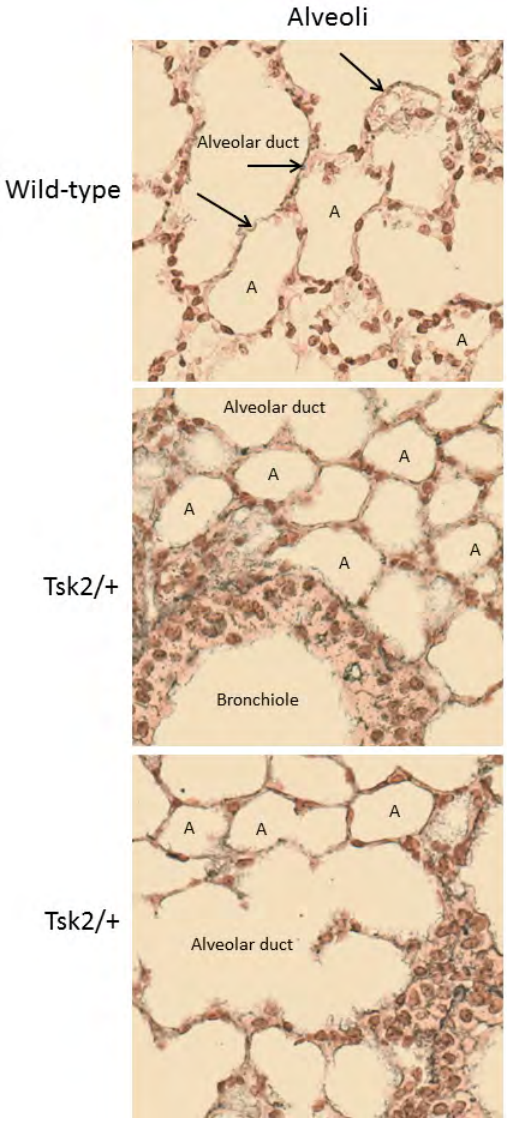


Figure 8. Tsk2/+ alveoli ("A") have diffuse and disorganized reticular fibers (gray stain), whereas there was well defined collagen in the wild-type mouse (arrows).

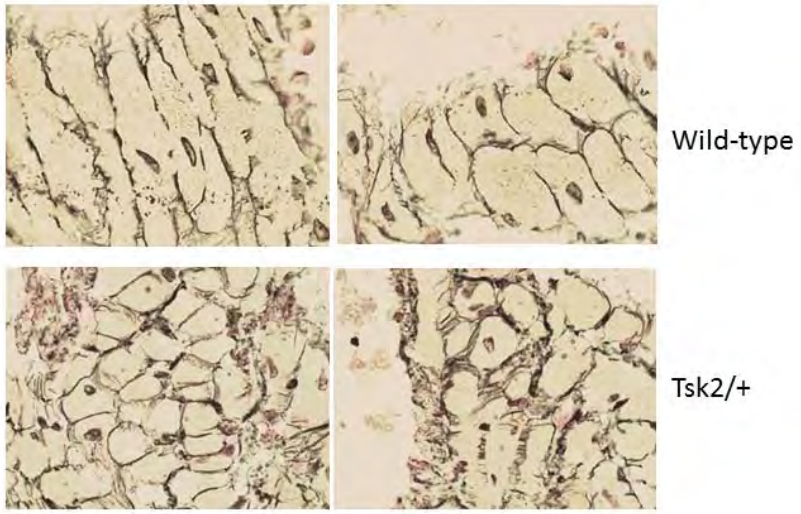


Figure 9. Tsk2/+ mice demonstrate increased reticular fibers surrounding epithelial cells of the lung. Note the increased deposition of stain between the cells in the Tsk2/+ lung.

KEY RESEARCH ACCOMPLISHMENTS (July 1 2012-June 30, 2013)

- We have completed the nucleotide sequencing of the mouse Tsk2/+ interval. A number of single nucleotide polymorphisms were found and verified by SNP allele-specific PCR analysis (ref 1) for these previously unsequenced strains (101/H, the parental strain; and B6.tsk2, the B6 congenic strain bearing mutated 101/H DNA on chromosome 1 that contains the Tsk2/+ mutation).
- *Col3A1* is the Tsk2 gene. This SNP is the only one in the interval that shows a coding mutation; it also is interesting that it often shows an expression difference, unlike any of the other genes with non-coding SNPs. The SNP changes a cysteine at amino acid 33 to serine.
- The genetic complementation test of *Col3a1* (mice bred from the Tsk2/+ line to BALB.Col3A1KO mice) was unequivocal and proved *Col3a1* is the *Tsk2* gene
- We have continued to show that the excess deposition of collagen matrix does not occur until well after the Tsk2/+ tight skin phenotype is evident.
- We have prepared a large number of fibroblast cell lines in preparation for *in vitro* studies of the Tsk2 trait under selective conditions.
- Microarray profiling of Tsk2/+ mice at 4, 8, 12 and 20 weeks have now been performed for both male and female mice totaling > 48 samples (we have previously only reported results at two time points for a single gender that totaled 8 samples). We have identified genes that are differentially regulated at each time point. Our prior data was suggestive of increased TGFβ signaling but did not show this directly using an experimentally derived TGFβ gene signature. We have now demonstrated that this is indeed the case. Tsk2/+ mice have a TGFβ1 signature that is seen in a global assessment of mRNA from skin in a carefully controlled study using littermates (Tsk2/+ and WT) at timed stages and stratified by sex. Controlling for hair cycle dramatically improves our ability to detect differently expressed genes and we are now mining these carefully controlled data for differentially expressed pathways.
- We found that the lung tissue of the Tsk2 mice has abnormal reticular fibers and this supports our findings of abnormal reticular fibers in the skin.

REPORTABLE OUTCOMES (July 1, 2012 to June 30, 2013)

Abstracts and Presentations:

1. C.M. Burgwin, K.B. Long, Zhenghui Li, C.M. Artlett, M. Whitfield, and E.P. Blankenhorn. 'Mapping the Mutation in the Tight Skin 2 model of systemic sclerosis' Drexel University College of Medicine Discovery Day Poster (October 2012).
2. Zhenghui Li, Eleni Marmarelis, Qu Kun, Lionel Brooks, Gavin D. Grant, Patricia A. Pioli, Howard Chang, Robert Lafyatis, and Michael L. Whitfield. "RNA-seq and miR-seq analysis of SSc skin across intrinsic gene expression subsets shows differential expression of non-*coding RNAs regulating SSc gene expression". Functional Genomics Data Society 15th Annual Meeting, Seattle, Washington, June 2013*
3. Zhenghui Li, Kristen Long, Carol Artlett, Elizabeth Blankenhorn and Michael L. Whitfield. "Identifying Genetic Variations in the Tsk2/+ Model of Scleroderma". Dartmouth MCB Program Annual Retreat. Whitefield, New Hampshire, August 2012

Manuscripts:

1. Long, K.B., Li, Z.., Burgwin, C., John, A.K., Sassi-Gaha, S., Earl, J., Eutsey, R., Ahmed, A., Ehrlich, G.D., Artlett, C.M., Whitfield, M.L., and Blankenhorn, E. P. The Tsk2/+ fibrotic phenotype is due to a gain-of-

function mutation in the PIIINP segment of the Col3a1 gene. PLOS Genetics, submitted, 2013 (in revision)

2. Long, K.B., Artlett, C.M., and Blankenhorn E. P. Tight Skin 2 Mice exhibit a novel time line of events leading to increased extracellular matrix deposition and dermal fibrosis, Matrix Biology, in preparation, 2013.
3. Long, K.B., Burgwin, C.M., Stairiker, C., Artlett, C.M., and Blankenhorn, E. P. The wound healing deficit in Tsk2/+ is due to excess elastin. Advances in Wound Care, in preparation, 2013.
4. Zhenghui Li, Kristen B. Long, Chelsea Burgwin, Carol M. Artlett and Elizabeth P. Blankenhorn, Michael L. Whitfield. The Tsk2/+ mouse model of Systemic Sclerosis shows variable age and gender dependent gene expression in skin. Manuscript *in preparation*.

Presentations:

6/13 Zhenghui Li (Whitfield lab), "RNA-seq and miR-seq analysis of SSc skin across intrinsic gene expression subsets shows differential expression of non-coding RNAs regulating SSc gene expression" FGED Meeting, Seattle WA.

6/13 Michael L. Whitfield, PhD "P30 SSc Microarray and Bioinformatics Core: Services to the national SSc community". Boston MA

5/13 Michael L. Whitfield, PhD "Personalized Medicine and Disease Pathogenesis in Systemic Sclerosis by Integrative Genomics" Geisel School of Medicine, Department of Micro-Immuno Annual Retreat.

4/13 Michael L. Whitfield, PhD "P50 Core: A focus on Next Generation Generation sequencing methods in SSc-PAH". Scleroderma NIAMS Center of Research Translation Advisory Committee and Investigators Meeting, Boston University School of Medicine, Boston MA.

3/13 Zhenghui Li (Whitfield Lab). “Identifying Genetic Variation in the Tsk2/+ Model of Systemic Sclerosis” Molecular and Cellular Biology Program, Research in Progress.

3/13 Michael Whitfield, PhD. “Plasticity and interconnectivity of the SSc intrinsic Subsets” Scleroderma Research Foundation Workshop, San Francisco, CA

3/13 Michael Whitfield, PhD. “Integrative Genomics in Systemic Sclerosis”, Stanford Epithelial Biology Program. Stanford, CA

11/12 Michael Whitfield, PhD. “NIAMS p30 Microarray Core for Scleroderma”, P30 Investigators and Users National presentation. Boston, MA

10/12 Michael Whitfield, PhD. “Subsetting systemic sclerosis by high-throughput gene expression”. Scleroderma SCOT clinical trial investigators meeting (NIH sponsored), Potomac MD

9/12 Chelsea Burgwin, year 3 graduate student supported by DOD. Molecular Cell Biology and Genetics Seminar Series, Drexel University College of Medicine (September 2012): "Mapping the Mutation in the Tight Skin 2 model of systemic sclerosis and Development of an in vitro Model."

8/12 Elizabeth Blankenhorn, " Genetic control of fibrosis in the Tsk2 Model of Scleroderma". Whitfield Laboratory Annual Retreat. New Hampshire, August 2012

6/12 Michael Whitfield, PhD. “Capturing the heterogeneity in systemic sclerosis with high-throughput gene expression profiling” Scleroderma NIAMS Center of Research Translation Advisory Committee and Investigators Meeting, Boston University School of Medicine, Boston MA

6/12 Michael Whitfield, PhD. “Intrinsic Gene Expression Subsets of Systemic Sclerosis Are Stable in Serial Skin Biopsies and Show Differential Therapeutic Response” Division of Rheumatology, Brigham and Women’s Hospital, Harvard Medical School, Boston MA.

6/12 Michael L. Whitfield, PhD. “Capturing the heterogeneity in systemic sclerosis with high-throughput gene expression profiling”. Synergy Translation Research Program, Geisel School of Medicine at Dartmouth. Lebanon, NH

Degrees obtained that are supported by this award

Chelsea M. Burgwin is working on her PhD thesis, expected defense in Spring, 2015.

Zhenghui Li is working toward his PhD thesis with an expected completion data in 2015.

Development of cell lines, tissue or serum repositories

We have developed a large number of Tsk2/+ and WT littermate fibroblast cell lines from mice of various ages and both sexes.

CONCLUSION:

A non-synonymous mutation in COL3A1 is almost certainly the causative mutation in Tsk2/+. This mutation results in a time dependent activation of numerous gene expression pathways, including TGFβ signaling, that varies by gender and age. The mechanism by which the COL3A1 mutation causes these changes is yet unknown but something we will be examining carefully in future studies.

REFERENCES

1. Bunce, M. *et al.* Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens* 46, 355-367 (1995).
2. Ishigame, H., Mosaheb, M.M., Sanjabi, S. & Flavell, R.A. Truncated Form of TGF-betaRII, But Not Its Absence, Induces Memory CD8+ T Cell Expansion and Lymphoproliferative Disorder in Mice. *Journal of immunology* 190, 6340-6350 (2013).
3. Sargent, J.L. *et al.* A TGF beta-Responsive Gene Signature Is Associated with a Subset of Diffuse Scleroderma with Increased Disease Severity. *J. Invest. Dermatol.* 130, 694-705 (2010).
4. Peters, J. & Ball, S.T. Tight skin-2 (Tsk2). *Mouse News Letter* 74, 91-92 (1986).
5. Oleggini, R., Gastaldo, N. & Di Donato, A. Regulation of elastin promoter by lysyl oxidase and growth factors: cross control of lysyl oxidase on TGF-beta1 effects. *Matrix biology : journal of the International Society for Matrix Biology* 26, 494-505 (2007).
6. Quaglino, D., Jr. *et al.* Connective tissue in skin biopsies from patients suffering systemic sclerosis. *J Submicrosc Cytol Pathol* 28, 287-296 (1996).

APPENDIX

1. Manuscript in revision for PLOS-Genetics, 2013.

The Tsk2/+ mouse fibrotic phenotype is due to a gain-of-function mutation in the PIIINP segment of the *Col3a1* gene.

Kristen B. Long[1], Zhenghui Li[2], Chelsea Burgwin[1], Alan K. John[1], Sihem Sassi-Gaha[1], Josh Earl[3], Rory Eutsey[3], Azad Ahmed[3], Garth D. Ehrlich[3], Carol M. Artlett[1], Michael L. Whitfield[2], and Elizabeth P. Blankenhorn[1, 4]

1, Department of Microbiology and Immunology, Drexel University College of Medicine, 2900 Queen Lane, Philadelphia, PA 19129

2, Geisel School of Medicine at Dartmouth, Department of Genetics, HB 7400, Hanover NH 03755

3, Center for Genomic Sciences, 320 East North Ave, Pittsburgh, PA 15212

4, Corresponding author

Elizabeth P. Blankenhorn

Elizabeth.Blankenhorn@drexelmed.edu

Phone: 215-991-8392

Fax: 215-848-2271

Abstract

Systemic sclerosis (SSc) is a polygenic, autoimmune disorder of unknown etiology, characterized by the excessive accumulation of extracellular matrix (ECM) proteins, vascular alterations, and production of autoantibodies. Among the major genetic models of SSc are the Tight skin (Tsk) 1/+ mouse and the Tsk2/+ mouse. The cause of fibrosis in Tsk1/+ mice is a duplication in the fibrillin-1 gene, and while this model has been thoroughly characterized, the disease signs in the Tsk1/+ mouse vary from SSc symptoms. The Tsk2/+ mouse model has many features of SSc including tight skin, excessive deposition of dermal ECM proteins, and the occurrence of autoantibodies, yet the gene causing this SSc-like disease is unknown. Through a series of backcrossing and collection of recombinant mice, we mapped the *Tsk2* gene mutation to an interval spanning less than 3 megabases (Mb) on chromosome one. We performed both RNA sequencing of skin RNA transcripts and genome capture DNA sequencing of the reduced genomic interval in tight-skinned and wild-type littermates to discover mutation(s) that could cause the SSc-like disease in Tsk2/+ mice. Multiple single and dinucleotide changes were observed, and all but three were ruled out by subsequent analyses. A missense point mutation in *Col3A1* remained as the best candidate for *Tsk2*, so a genetic complementation test was used to prove that this *Col3a1* mutation causes skin fibrosis in the Tsk2/+ mice. All previously documented mutations in the *Col3a1* gene are associated with Ehlers-Danlos syndrome, a connective tissue disorder that leads to a defect in type III collagen synthesis. The *Tsk2* point mutation, localized to the procollagen III amino terminal propeptide segment of COL3A1, is the first documented gain-of-function mutation associated with *Col3a1* and it leads instead to fibrosis. This discovery provides insight into a new mechanism of fibrosis manifest by the Tsk2/+ mice.

Author Summary (non-technical summary):

SSc is a rare, potentially fatal autoimmune disease that causes skin and internal organ fibrosis. There is no cure for the disease and only palliative treatment is available. One of the factors limiting the treatment of SSc is a lack of understanding of underlying disease mechanisms, and although genetic and environmental factors seem to be important for disease development, the specific triggers are unclear. It is therefore necessary to uncover the early changes that cause fibrosis in order to develop new treatments for inhibiting disease progression. Skin fibroblasts and biopsy samples from patients can only be investigated after disease has been well established, highlighting the need for an animal model that will allow for the investigation of early skin changes before the onset of fibrosis. We use the Tight-skin2 model (Tsk2/+). Here, we report that a mutation in the *Col3a1* gene is responsible for disease in the Tsk2/+ mouse. This is the first time that a *Col3a1* mutation has been linked to excessive collagen accumulation and presents as an exciting new gene to study the development of fibrosis.

Introduction

There are multiple animal models of SSc [1], yet none mimics all facets of SSc disease. Of the genetic models, the cause of disease in tight-skin 1 (Tsk1/+) mice is known to be a tandem duplication in the *Fbn1* gene [2]. The Tsk1/+ mouse has been useful for studying SSc, but disease signs differ somewhat from human SSc symptoms. Tsk1/+ mice develop excess skin fibrosis; however the collagen accumulation occurs in the hypodermis [3], as compared with dermal accumulation of collagen seen in SSc [4]. Lung disease in Tsk1/+ mice mimics an emphysema-like disease [5], whereas involved lung tissue in patients with SSc is characterized by fibrosis that can also be complicated by vasculopathy and pulmonary arterial hypertension [4,6]. Additionally, a limited number of Tsk1/+ mice develop antinuclear antibodies (ANAs) [3], compared to over 90% of SSc patients with ANAs [7].

Other models of SSc have employed mice with individual gene deficiencies or overexpression (for example, Fos-related antigen-2 [8], endothelin [9,10] and Friend leukemia integration factor-1 [11]) have proven useful for understanding the contribution of these proteins to the vasculopathy and/or lung fibrosis seen in scleroderma. Non-genetic models of SSc include the bleomycin model [12], which has been used to study many of the initiating events involved in fibrosis after bleomycin is instilled into the lung or injected under the skin. Indeed, we have recently shown that this model exhibits fibrosis that is dependent on the activation of the innate immune system mediated by the NLRP3 inflammasome to drive the fibrotic response to bleomycin [13]. In this model, vascular remodeling does occur [14] and the fibrotic effects can be systemic [1,15]; however, ANAs are not always observed, and the signs of disease last for only ~6 weeks after the end of treatment.

The Tsk2/+ mouse was first described in 1986, when an offspring of a 101/H mouse exposed to the mutagenic agent ethylnitrosourea was noted to have tight skin in the interscapular region [16]. The gene causing SSc-like signs in Tsk2/+ mice was reported to be located on chromosome 1 between 42.5 and 52.5 Mb [17], however the genetic defect was never identified. Like *Tsk1*, *Tsk2* SSc-like traits are highly penetrant in Tsk2/+ heterozygotes and like Tsk1/+ it is homozygous embryonic lethal. Tsk2/+ mice have many features of human disease including tight skin, dysregulated dermal extracellular matrix (ECM) deposition, and significant autoimmune response [18,19], making this mouse ideal for use as one model of SSc.

Currently, there are about 20 known genes and multiple predicted genes in the *Tsk2* region. Of the known 20 genes, there are three main candidate genes: transforming growth factor beta receptor associated protein 1 (*Tgfbrap1*), collagen, type III, alpha 1 (*Col3a1*), and collagen, type V, alpha 2 (*Col5a2*). Herein, we report the identity of the *Tsk2* gene. We have discovered a gain-of-function missense mutation in *Col3a1,* and we hypothesize that it causes the altered COL3A1 protein to directly influence TGF-β1 signaling, an important mediator of collagen production. This mouse affords a unique opportunity to examine the pathways leading to the multiple clinical parameters of fibrotic disease from birth onward.

Results

Our investigations on the identity of the *Tsk2* gene were initiated with further mapping of the *Tsk2* interval by genotyping backcross progeny of Tsk2/+ mice bred to C57Bl/6 (B6) mice. We screened phenotyped littermate mice for informative microsatellites: *D1Mit233*, *D1Mit235*, a microsatellite in *Gls,* and *D1Mit18*. Multiple recombinants were recovered that together exhausted the current microsatellite boundaries. Because there were no more informative microsatellites in the interval, the interval was fine-mapped using SNP typing between 42.53 and 52.22 Mb on chromosome 1. 101/H and B6 strains are extremely similar genetically on chromosome 1, and only one SNP differentiated 101/H from B6, making SNP typing less useful to perform linkage mapping. Therefore, each recombinant was bred and then backcrossed to a consomic B6.chr 1-A/J mouse to fine-map the region by SNP typing, as A/J mice bear many known SNPs compared to B6 mice. Additional recombinants were recovered, narrowing the *Tsk2* interval to between 44.67 – 47.11 Mb (**Fig. 1A**). With this approach we narrowed the interval to less than 2.5 Mb, representing a greater than 3-fold reduction of the size of the interval bearing 101/H genomic DNA and *Tsk2*. There are seven known genes in this interval (**Fig. 1B**).

To identify the mutation underlying *Tsk2*, we employed both RNA sequencing (RNA-Seq) and genome capture sequencing of the reduced genomic interval. For RNA-Seq, total RNA was prepared from skin and sequenced from 4 Tsk2/+ and 3 wild type (WT) littermates. We also sequenced total RNA from liver of the 101/H parental mouse strain of Tsk2/+ mice. Reads were aligned to the MM9 reference genome (B6) and analyzed for polymorphisms in the *Tsk2* interval. We restricted our analysis to genetic variants found in all four Tsk2/+ mice to eliminate sequencing errors. Among these were 265 single nucleotide polymorphisms (SNPs) found in both WT and Tsk2/+ littermates that represent differences between the reference B6 genome and the 101/H background, so these were excluded from further study. Thirteen SNPs were found in all four Tsk2/+ mice analyzed; ten of these SNPs were also found to be in liver RNA from 101/H strain or in other non-fibrotic mouse strains (http://phenome.jax.org/), and thus were also ruled out as candidates for *Tsk2* (**Table 1**). The remaining three SNPs were heterozygous and confirmed to be only in Tsk2/+ mice. One of these, in a *Gulp1* intron, proved useful as a new marker that resides **outside** the supported interval for Tsk2/+ in an informative recombinant mouse (**Fig. 1A**). The second SNP was also found in an intron of *Gulp1*. Analysis of the RNA-Seq data did not find any splicing defects in *Gulp1* mRNA in the Tsk2/+ mice (**Supplementary Fig. 1**), indicating that this SNP does not change mRNA splicing, and its gene expression in skin is unchanged (**Fig. 2**). Thus the intronic SNP in *Gulp1*

is unlikely to play a role in the tight skin phenotype. The final SNP is in *Col3a1*. The mutation was confirmed by resequencing the Tsk2/+ *Col3a1* gene, including both 5' and 3' UTRs, whereby we verified that the SNP in *Col3a1* results in a T to A transversion at Chr1:45,378,353, causing a Cys → Ser amino acid change in the procollagen III amino terminal propeptide (PIIINP) segment, a natural cleavage product of COL3A1. The mutant protein is designated COL3A1^{Tsk2} (C33S). No other gene in the *Tsk2* interval except *Col3a1* is highly expressed in skin, and RNA-Seq results indicate that this mutation leads to an increase of *Col3a1* transcripts in 4-week old Tsk2/+ skin samples compared to WT littermates (**Figure 2**).

Because RNA-Seq only captures variation in the transcribed regions of the genome, and thus might miss an important genomic feature that is unique to *Tsk2*, we sequenced captured genomic DNA samples corresponding to the minimal linkage region from B6.Tsk2/+ heterozygotes and 101/H homozygous parental strain mice. The Tsk2/+ and 101/H DNA samples were first selected on Nimblegen arrays representing an ~3 Mb interval containing *Tsk2*, and then were submitted to 454 sequencing. Alignment revealed multiple DNA differences between the Tsk2/+ mouse and its parental 101/H strain. A majority of the differences observed were accounted for by non-chromosome 1 repetitive DNA sequences such as LINE, SINE and retroviral elements captured by the Nimblegen array due to homology with repetitive sequences contained within the *Tsk2* interval on chromosome 1. After filtering these repetitive elements from the comparison, there were six single copy DNA sequence differences, of which three were confirmed to be Tsk2/+ specific (**Table 1**). Among these, the T-to-A transversion in *Col3A1* at Chr1:45,378,353, identical to the mutation identified by RNA-Seq, was the most likely to be *Tsk2* by genomic assessment.

To prove that *Tsk2* is a single nucleotide change in the *Col3A1* coding region required a separate genetic test. Both *Tsk2/Tsk2* [16] and *Col3a1*-KO [20] homozygotes exhibit embryonic lethality, which is also seen in our mouse colony **(Table 2).** We therefore designed a complementation study to determine if *Col3a1* is identical to the gene underlying Tsk2/+ pathology. Tsk2/+ x *Col3a1*$^{-/+}$ mice were bred together, and 33 progeny mice were genotyped. No viable compound heterozygotes were born (**Table 2)**, indicating that the *Tsk2* mutation cannot be rescued by any other gene if *Col3a1* is null on the other chromosome. This result also proves that there was no compensation by a putatively normal COL3A1 protein from the Tsk2-origin chromosome, which would have resulted in normal mouse development and survival *in utero*. In fact, having the *Tsk2* mutation is more damaging than not expressing COL3A1 at all, because while a

few percent of *Col3a1*-KO homozygotes make it to birth, *Tsk2/Tsk2* homozygotes never do, and whereas *Col3a1*^{Tsk2} mice are viable but small in stature and fibrotic, *Col3a1*$^{-/+}$ heterozygotes are normal. We conclude that the mutation in Tsk2/+ mice lies within *Col3a1* and is substantially more deleterious than a complete genetic deficiency of COL3A1.

COL3A1 is the second most abundant collagen in the skin after type I collagen, making up about 10-15% of the total collagen[21]. COL3A1 production is important during embryogenesis[22] and in wound healing responses as it is expressed in early granulation tissue[20,23,24]. SSc traits have been compared to an uncontrolled wound healing response[21], and fibroblasts from SSc patients have been found to produce increased levels of *Col3a1* message[25]. The behavior of *Col3a1* in Tsk2/+ mice could reveal the mechanism by which this mutation causes very substantial ECM fibrosis and very tight skin.

To determine if there is excess production of COL3A1 protein in Tsk2/+ mice, and if the mutation is cell-autonomous, fibroblasts were grown from skin explants of Tsk2/+ and WT littermates at birth. Protein production was analyzed by hydroxyproline analysis for all skin collagens and western blot analysis for COL3A1. Tsk2/+ fibroblasts produce significantly more collagen protein (not shown) and specifically more COL3A1 compared to WT littermate fibroblasts (**Figure 3**). We conclude the *Tsk2* mutation is cell-autonomous in that it does not require an intact setting *in vivo* to demonstrate a significant increase in expression of COL3A1 protein.

We localized the excess production of COL3A1 protein by histological examination of Tsk2/+ and WT littermate skin. Reticular fibers are composed primarily of COL3A1 and are a structural element in the skin, found in the panniculus carnosus and in the dermis. COL3A1 expression in skin from two-week old mice is high and declines after birth in WT littermates, but does not decline in the Tsk2/+ mice (**Fig. 4**). As Tsk2/+ mice age, the reticular fibers thicken and become more pronounced compared to their WT littermates reflecting the increase in COL3A1 deposition.

Discussion

Sequencing of both rRNA depleted expressed RNAs and the genomic region in the Tsk2/+ interval, coupled with the genetic complementation study, prove that Tsk2/+ mice harbor a deleterious coding mutation in *Col3a1*, leading to an amino acid change (C33S) in the N-terminal region of the protein (PIIINP). Effects of the *Tsk2* mutation include: 1) higher COL3A1 protein expression in fibroblast cell lines and *in vivo*; 2) a more lethal phenotype than the homozygous genetic loss of *Col3a1*; and 3) a more lethal compound heterozygous phenotype than that of the homozygous gene knockout, indicating that COL3A1^{Tsk2} C33S has a dominant prenatal lethal effect. Of interest is that this mutation not only changes the amino acid sequence, but also its expression levels: *Col3a1* transcripts show increased expression in Tsk2/+ mice, compared to WT littermates of the same sex and age (Long, et al., in preparation, 2013). This change in expression is age-dependent and appears to be different in female mice *vs.* male mice.

This is the first mutation in *Col3a1* that results in a gain-of-function phenotype instead of Ehlers-Danlos-like syndromes that are due to loss-of-function or antimorphic collagen-poor phenotypes. Ehlers-Danlos is a group of connective tissue disorders characterized by highly elastic, fragile but not fibrotic skin due to a defect in collagen synthesis [26]. In addition, these patients have a significant risk for aneurism. Ehlers-Danlos syndrome has been associated with 337 mutations in COL3A1 (http://www.le.ac.uk/ge/collagen/), in addition to mutations on COL1A1 and COL5A2. These mutations result in amino acid substitutions, RNA splicing alterations, deletions, or null alleles. Currently, all reported COL3A1 mutations result in decreased collagen protein secretion leading to thinner skin and defects in the vasculature that are observed in these patients. In contrast to the mutations observed in Ehlers-Danlos, the Tsk2/+ mouse mutation results in thickened skin with no apparent evidence of aneurism.

Once type III collagen is secreted from the cell, there is additional extracellular processing that cleaves off the amino terminal propeptide sequence, PIIINP, by N-proteinase [27]. The PIIINP molecule comprises three domains: a cysteine-rich globular domain (Col 1) containing 79 amino acids with five intrachain disulphide bonds, a triple-helical domain (Col 3) with 12 amino acids and three interchain disulphide bonds, and a non-collagenous domain (Col 2) comprising of 39 amino acids ending with the N-telopeptide that forms a triple helical structure [28]. PIIINP has a molecular weight of approximately 42,000 daltons. The function of PIIINP is thought to regulate fibrillogenesis of type I

collagen resulting in thinner collagen fibrils [29]. In the absence of PIIINP, Romanic and colleagues demonstrate that COL1A1 is larger, shorter, and apparently stiffer; whereas in the presence of PIIINP/COL1A1 copolymer, the type I collagen was longer, thinner, and more flexible. The mutation identified in *Col3a1* in the Tsk2/+ mouse localizes to the Col 1 domain in close proximity to the transglutamination site. We speculate that this mutation may interfere with the copolymerization between collagen I and the PIIINP molecule resulting in thicker collagen fibers. The PIIINP molecule may also be involved in the fibrillogenesis of type III collagen as we observed thicker reticular fibers as early as 2 weeks of age (Figure 4). In addition, the transglutamination site has been found to play a role in binding of other molecules such as fibronectin [30] and may also bind fibrinogen [31]. Alterations near this site could change skin homeostasis and cellular signaling leading to fibrosis. Current studies are underway to test the hypothesis that the function of PIIINP is altered in Tsk2/+ mice, leading to fibrosis.

Furthermore, because heterozygous Tsk2/+ mice have a pronounced fibrotic phenotype, we also hypothesize that the mechanism by which *Tsk2* causes disease and alters the normal ECM involves heterodimers of mutant and WT collagen strands. This hypothesis is attractive: it would likely be dominant within the heterozygote, as collagen III is a homotrimeric triple helix [32], and the gene product of the mutant chromosome could be expected to contribute to altered folding of a majority of collagen helices even in the presence of 50% normal collagen [33]. The mechanism for the severe fibrotic effects of *Tsk2* is currently under investigation in our laboratories, based on these hypotheses.

Materials and Methods

Ethics Statement: All studies and procedures were approved by the Institutional Animal Care and Use Committee at Drexel University College of Medicine, and conducted in accord with recommendations in the "Guide for the Care and Use of Laboratory Animals" (Institute of Laboratory Animal Resources, National Research Council, National Academy of Sciences).

Mice: Breeding pairs of Tsk2/+ mice were obtained from Dr. Paul Christner at Jefferson University and housed at Drexel University College of Medicine. The Tsk2/+ mice (originating on a 101/H background) were bred initially onto a hybrid (C3H x C57Bl/6J (B6)) background. Serial backcrossing of Tsk2/+ mice to B6 mice was initiated, and the mice in this study were backcrossed at least 5 generations. Recombinant B6.Tsk2/+ mice were bred to B6.chr 1-A/J mice (Jackson Laboratory, Bar Harbor, ME) and the resulting B6.Tsk2/+ F1 mice were backcrossed to B6.chr 1-A/J mice to fine-

map *Tsk2*.

DNA isolation: Tail snips were taken from weanlings, and DNA was isolated from the tissue using a GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO), following the manufacturer's protocols. DNA was diluted to a working concentration of 40 ng/µl.

Microsatellite typing: 120 ng of DNA per reaction was used when amplifying microsatellites by polymerase chain reaction (PCR). Microsatellite *D1Mit233*, sequence forward 5'-TAGACCCATCACTTTCCAAG-3' and reverse 5'-ACTGGCTAAAGTA TCCTAGAAAGGG-3' was run at an annealing temperature of 49 ^{o}C. Microsatellites *D1Mit235* sequence forward 5'-CACCTGGCTAAGAGACCATACC-3' and reverse 5'-GCCTCCACTACCACCATCTC-3'; a microsatellite in the *Glutaminase* gene (*Gls*) sequence forward 5'-TGTGCACTT GAGAATTTTGCTT-3' and reverse 5'-CCCACATACTGGACCTACCC-3'; and *D1Mit18* sequence forward 5'-TCTGGTTCCAGGCTTGATTC-3'and reverse 5'-TCACAAGTGA GGCTCCAGG-3' were run at an annealing temperature of 50 ^{o}C. PCR products were separated by electrophoresis on a 3% agarose gel.

SNP typing: SNP typing as previously described [34] was used to map the boundaries of the recombinations. Specific locations of polymorphisms between B6 (which is very similar to 101/H) and A/J were determined using Mouse Genome Informatics (www.informatics.jax.org), and primers for SNP typing were designed using Primer 3 online software (http://frodo.wi.mit.edu) and synthesized by Integrated DNA Technologies (www.idtdna.com). 100 ng of DNA per reaction was used when amplifying SNPs by PCR. PCR products were separated by electrophoresis on a 1.5% agarose gel.

RNA-Seq – Sample RNA preparation: Total RNA was prepared from three WT and four Tsk2/+ mice skin biopsies using Qiagen RNeasy Fibrous Tissue Mini Kit. The RNA integrity (RIN) was determined by Agilent Bioanalyzer Nano chip. All samples used for this study had RIN scores of 7 or greater [35].

RNA-Seq: RNA-seq sequencing libraries were prepared for the 7 samples (3 WT and 4 Tsk2/+) using NuGEN Ovation RNA-Seq System (NuGen, San Carlos, CA). Libraries were multiplexed and sequenced on an Illumina HiSeq 2000 platform to obtain 16.7-50.9 million 50 bp paired-end reads per sample. The raw reads were aligned to the reference mouse genome (MM9 assembly) using Tophat software with default parameters [36,37]. Supplemental Figure 1 shows RNA-Seq read coverage for three interval genes.

Variant calling and SNP identification: In order to identify coding region and intronic genetic variation we analyzed the data using Freebayes software [35]. The following parameters were used: "freebayes –r 1:44200000..47100000 –f MM9.fai –b <tophat_aligned_bamfile> -v <output_SNP.vcf>" We compared the WT and Tsk2/+ mice to the MM9 reference genome strain (B6).

454 Sequencing: Sequence capture array probes were designed by Roche Nimblegen using the mouse genome sequence between 44,241,286 and 47,116,890 on chromosome 1 of mouse genome (mm9). Probes were designed corresponding to 56.3% of the linkage region, however in practice a larger area was captured due to the overhang of the larger fragments (~600bp) being used. Probes could not be designed to the remaining 43.7% due to it being composed of repetitive sequence. Samples were captured and amplified as described in the Roche Nimblegen sequence capture manual (Version 1.0). Based on qPCR analysis, control capture regions showed an enrichment range of 14 to 50 fold for the samples tested. Titanium general libraries were prepared from the captured DNAs from two 101/H mice (samples 101B S. Wells and 101E) and two Tsk2/+ mice (samples 2044 and 2045) using 5000 ng of DNA as described in the GS FLX Titanium, General Library Preparation Method Manual, October, 2008 (Roche Molecular Systems, Nutley, NJ. GS FLX multiplex identifiers (MID1, MID2, MID3 and MID4) were used during the library preparation. Enriched captured fragments binding to beads, titration, emulsion PCR, emulsion breaking, bead enrichment, and pico-titer plate-based pyrosequencing were performed as described in GS FLX Titanium emPCR and Sequencing Protocols, October, 2008. A total of 6 million beads (1,500,000 beads for each of the four mice) were used for the sequencing which consumed one and one -half pico-titer plates.

Multiplexed 454 sequenced reads were assembled using Newbler v2.6 with scaffolding against the same chromosome region that the probes were derived from. Separate assemblies were created for each of the four mice by MID number, and lists of variants for each mouse were obtained from the assembler output. Variants were filtered by quality (phred scores >30), depth of coverage (>13 reads), and heterozygousity (>20% of reads differed from the reference). Variants were mapped to exons , introns, and intergenic regions within the linkage region and set analysis between the 204 and 101 lineages were performed using custom perl scripts. Sets were examined for variants between the 204 and 101 line, and between all samples and the reference to identify heterozygous SNPs uniquely present in the tsk2 line.

Complementation analysis with Col3a1$^{-/+}$ mice: Col3a1$^{-/+}$ mice (bearing one allele containing *Col3a1* and one allele where *Col3a1* has been knocked out) were received as a generous gift from Xianhua Piao at Harvard Medical School. Tsk2/+ mice were crossed to Col3a1$^{-/+}$ mice to verify that the SNP in *Col3a1* is *Tsk2*, as Tsk2/Tsk2 homozygous mice are not viable, and if *Tsk2* is located in *Col3a1*, then Tsk2.Col3a1^{-} mice will not be viable. The resulting first generation of the cross was genotyped by PCR for Tsk2/+ using the microsatellites listed above and primers specific to *Col3a1* or the inserted neomycin cassette. WT forward (common) primer 5'-CTTCTCACCCTTCTTCATCCC-3', WT reverse primer 5'-AGCCTGTTCAATCGGTACC-3', and neomycin reverse primer 5'- GCTATCAGGACATAGCGTTGG-3'. A second primer set was used to verify the knock out, WT forward (common) primer 5'-AGGGCCTTCAGAGGATTTTC-3', WT reverse primer 5'-CCATCCCCTCAGCAGTAAA-3', and the neomycin reverse 5'-GFCCAGAGGCCACTTGTGTAG-3'. Reactions were run at an annealing temperature of 58 ^{o}C and PCR products were separated by electrophoresis on a 1.5% agarose gel.

***In vitro* fibroblast culturing:** Fibroblasts were grown from skin explants of Tsk2/+ and WT neonatal mice, approximately 1 day old. Briefly, neonates were sacrificed and the skin was cleaned using 70% EtOH, carefully dissected from the body and placed in Dulbecco's Modified Eagle Medium (Invitrogen Life Technologies, Carlsbad, CA) with 20% FBS (Gemini Bioproducts, West Sacramento CA), 1% penicillin-streptomycin (Cellgro, Manassas VA) and 1% Glutamax (Gibco Life Technologies, Grand Island NY). The skin explants were minced into small squares and cultured in 6 cm^2 tissue culture dishes. The cells were kept in a 5% CO_2/95% air moist incubator at 37 ^{o}C and cultured until confluent. The tails were removed from the neonates for DNA isolation and Tsk2/+ microsatellite and sex genotyping by PCR.

Western blot analysis of COL3A1 expression: When confluent, collagen production *in vitro* was determined by western blot analysis. Briefly, 5x10^5 neonatal fibroblasts from Tsk2/+ and WT mice were cultured in 10 cm^2 dishes until confluent. Once confluent, culture supernatants were collected, cells were washed with PBS and then lysed on ice using RIPA buffer (Sigma-Aldrich, St Louis MO) containing 1% phosphatase inhibitor and 1% protease inhibitor (Sigma-Aldrich, St Louis MO) for 5 minutes. Cells were then scraped from the dishes and centrifuged at 8,000 x g for 10 minutes at 4 ^{o}C to pellet debris. The supernatant was collected and total protein was measured with a Bradford assay (Sigma-Aldrich, St Louis MO). Approximately 75 µg of protein per sample was added to reducing buffer, boiled, and then loaded onto an 8% SDS gel. After separation, proteins were transferred to a polyvinylidene fluoride membrane. The membrane was blocked in 5% nonfat milk in Tris buffered saline and then probed with goat anti-Col3a1 (#sc-8781, Santa Cruz

Biotechnology, Inc, Santa Cruz, CA) or rabbit anti-β-Actin (#4967, Cell Signaling Technologies, Boston, MA) and then probed with a secondary antibody, donkey anti-goat (#705-035-003, Jackson ImmunoResearch Laboratories, West Grove, PA) or goat anti-rabbit (#111-035-003, Jackson ImmunoResearch) respectively, and developed using SuperSignal West Dura ECL reagent (Thermo Scientific Inc, Rockford, IL). Band intensities were measured using ImageQuant TL Software (GE Healthcare Life Sciences).

Reticular fiber staining: Reticular fibers were stained using the Chandler's Precision Reticular Fiber Stain kit (American Master*Tech, Lodi CA) according to the manufacturer's protocol.

Statistics: For Figure 2, two-way ANOVA was used to determine statistical significance of gene expression by RNA-Seq for the 7 genes in the interval. One data point was removed for *Col3A1* expression group for a WT mouse, as it was determined to be an outlier by Grubbs' test at alpha=0.01. For Figure 3, a two-tailed student's t-test was used to determine statistical significance.

REFERENCES

1. Artlett CM (2010) Animal models of scleroderma: fresh insights. Curr Opin Rheumatol 22: 677-682.

2. Siracusa LD, McGrath R, Ma Q, Moskow JJ, Manne J, et al. (1996) A tandem duplication within the fibrillin 1 gene is associated with the mouse tight skin mutation. Genome Res 6: 300-313.

3. Baxter RM, Crowell TP, McCrann ME, Frew EM, Gardner H (2005) Analysis of the tight skin (Tsk1/+) mouse as a model for testing antifibrotic agents. Lab Invest 85: 1199-1209.

4. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, et al. (1988) Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 15: 202-205.

5. Szapiel SV, Fulmer JD, Hunninghake GW, Elson NA, Kawanami O, et al. (1981) Hereditary emphysema in the tight-skin (Tsk/+) mouse. Am Rev Respir Dis 123: 680-685.

6. Guiducci S, Giacomelli R, Cerinic MM (2007) Vascular complications of scleroderma. Autoimmun Rev 6: 520-523.

7. Okano Y (1996) Antinuclear antibody in systemic sclerosis (scleroderma). Rheum Dis Clin North Am 22: 709-735.

8. Maurer B, Busch N, Jungel A, Pileckyte M, Gay RE, et al. (2009) Transcription factor fos-related antigen-2 induces progressive peripheral vasculopathy in mice closely resembling human systemic sclerosis. Circulation 120: 2367-2376.

9. Richard V, Solans V, Favre J, Henry JP, Lallemand F, et al. (2008) Role of endogenous endothelin in endothelial dysfunction in murine model of systemic sclerosis: tight skin mice 1. Fundam Clin Pharmacol 22: 649-655.

10. Hocher B, Schwarz A, Fagan KA, Thone-Reineke C, El-Hag K, et al. (2000) Pulmonary fibrosis and chronic lung inflammation in ET-1 transgenic mice. Am J Respir Cell Mol Biol 23: 19-26.

11. Asano Y, Stawski L, Hant F, Highland K, Silver R, et al. (2010) Endothelial Fli1 deficiency impairs vascular homeostasis: a role in scleroderma vasculopathy. Am J Pathol 176: 1983-1998.

12. Yamamoto T, Takagawa S, Katayama I, Yamazaki K, Hamazaki Y, et al. (1999) Animal model of sclerotic skin. I: Local injections of bleomycin induce sclerotic skin mimicking scleroderma. J Invest Dermatol 112: 456-462.

13. Artlett CM (2012) The Role of the NLRP3 Inflammasome in Fibrosis. Open Rheumatol J 6: 80-86.

14. Yamamoto T, Katayama I (2011) Vascular changes in bleomycin-induced scleroderma. Int J Rheumatol 2011: 270938.

15. Artlett CM, Sassi-Gaha S, Rieger JL, Boesteanu AC, Feghali-Bostwick CA, et al. (2011) The inflammasome activating caspase 1 mediates fibrosis and myofibroblast differentiation in systemic sclerosis. Arthritis Rheum 63: 3563-3574.

16. Peters J, Ball ST (1986) Tight Skin 2 (Tsk2). Mouse News Letters: 91-92.

17. Christner PJ, Siracusa LD, Hawkins DF, McGrath R, Betz JK, et al. (1996) A high-resolution linkage map of the tight skin 2 (Tsk2) locus: a mouse model for scleroderma (SSc) and other cutaneous fibrotic diseases. Mamm Genome 7: 610-612.

18. Christner PJ, Peters J, Hawkins D, Siracusa LD, Jimenez SA (1995) The tight skin 2 mouse. An animal model of scleroderma displaying cutaneous fibrosis and mononuclear cell infiltration. Arthritis Rheum 38: 1791-1798.

19. Gentiletti J, McCloskey LJ, Artlett CM, Peters J, Jimenez SA, et al. (2005) Demonstration of autoimmunity in the tight skin-2 mouse: a model for scleroderma. J Immunol 175: 2418-2426.

20. Liu X, Wu H, Byrne M, Krane S, Jaenisch R (1997) Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. Proc Natl Acad Sci U S A 94: 1852-1856.

21. Jinnin M (2010) Mechanisms of skin fibrosis in systemic sclerosis. J Dermatol 37: 11-25.

22. Fessler JH, Fessler LI (1978) Biosynthesis of procollagen. Annu Rev Biochem 47: 129-162.

23. Hurme T, Kalimo H, Sandberg M, Lehto M, Vuorio E (1991) Localization of type I and III collagen and fibronectin production in injured gastrocnemius muscle. Lab Invest 64: 76-84.

24. Merkel JR, DiPaolo BR, Hallock GG, Rice DC (1988) Type I and type III collagen content of healing wounds in fetal and adult rats. Proc Soc Exp Biol Med 187: 493-497.

25. Zhou X, Tan FK, Guo X, Arnett FC (2006) Attenuation of collagen production with small interfering RNA of SPARC in cultured fibroblasts from the skin of patients with scleroderma. Arthritis Rheum 54: 2626-2631.

26. Nishiyama Y, Nejima J, Watanabe A, Kotani E, Sakai N, et al. (2001) Ehlers-Danlos syndrome type IV with a unique point mutation in COL3A1 and familial phenotype of myocardial infarction without organic coronary stenosis. J Intern Med 249: 103-108.

27. Halila R, Peltonen L (1986) Purification of human procollagen type III N-proteinase from placenta and preparation of antiserum. Biochem J 239: 47-52.

28. Bruckner P, Bachinger HP, Timpl R, Engel J (1978) Three conformationally distinct domains in the amino-terminal segment of type III procollagen and its rapid triple helix leads to and comes from coil transition. Eur J Biochem 90: 595-603.

29. Romanic AM, Adachi E, Kadler KE, Hojima Y, Prockop DJ (1991) Copolymerization of pNcollagen III and collagen I. pNcollagen III decreases the rate of incorporation of collagen I into fibrils, the amount of collagen I incorporated, and the diameter of the fibrils formed. J Biol Chem 266: 12703-12709.

30. Hedman K, Alitalo K, Lehtinen S, Timpl R, Vaheri A (1982) Deposition of an intermediate form of procollagen type III (pN-collagen) into fibrils in the matrix of amniotic epithelial cells. EMBO J 1: 47-52.

31. Bowness JM, Tarr AH, Wiebe RI (1989) Transglutaminase-catalysed cross-linking: a potential mechanism for the interaction of fibrinogen, low density lipoprotein and arterial type III procollagen. Thromb Res 54: 357-367.

32. Ramachandran GN, Kartha G (1955) Structure of collagen. Nature 176: 593-595.

33. Strachan T, Read AP (1999) Human Molecular Genetics.: Wiley-Liss, New York.

34. Bunce M, O'Neill CM, Barnardo MC, Krausa P, Browning MJ, et al. (1995) Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). Tissue Antigens 46: 355-367.

35. Erik Garrison GM (2012) Haplotype-based variant detection from short-read sequencing. http://arxivorg/abs/12073907.

36. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, et al. (2012) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol.

37. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7: 562-578.

Figure Legends

Figure 1. *Tsk2* lies between and not including 44.67 and 47.11 Mb on chromosome 1.

(A) The *Tsk2* interval was narrowed by screening back-crossed mice on the B6.chr 1-A/J background for microsatellites and SNPs. Grey bars (101/H) depict the original parental strain, bearing *Tsk2*. White bars depict the B6 genome. For clarity, indefinite areas between typed markers are not highlighted. Recombinants A – F are animals bearing novel recombination sites in the interval. The phenotypes are tight (T – Tsk2/+) and loose (L – WT). Together, these recombinants narrowed the interval to < 2.5 Mb.

(B) The *Tsk2* interval was narrowed to less than 3 Mb, therefore eliminating 13 genes. Currently, *Gulp1, Col3a1, Col5a2, Wdr75, Slc40a1, Dnahc7b* and *Slc30a10* remain in the interval. The elements of the *Gulp1* gene above 44.67 Mb are excluded from candidacy by the recombination in mouse F.

Figure 2. *Col3a1* is the only interval gene expressed at high levels in the skin of Tsk2/+ mice.

(A) This graph shows gene expression for the seven *Tsk2* interval genes, as determined from the RNA-Seq abundance results.

(B) *Col3a1* is the only gene of the seven to achieve significance by 2-way ANOVA. One data point for *Col3a1* expression (in a WT mouse) was removed, as it was determined to be an outlier by Grubbs' test at alpha=0.01. NS = not significant, *** = significant.

(C) Heat map for seven *Tsk2* interval genes detected as transcripts in RNA-Seq.

(D-E) Distribution of nucleotide calls in heterozygous Tsk2/+ and homozygous WT mice for *Col3a1* and *Gulp1*.

Figure 3. Fibroblasts derived from the skin of Tsk2/+ mice produce significant amounts of COL3A1 *in vitro.*

(A) Fibroblasts from Tsk2/+ neonates (1 day old) produce more collagen than fibroblasts from WT neonates. Fibroblasts were cultured from skin explants from Tsk2/+ and WT mice and evaluated for the production of collagen. COL3A1 protein levels in cell lysates were measured by western blot. (B) Western blot band densities were calculated using ImageQuant TL Software and normalized to the loading control β-actin (n=3)

Figure 4. Tsk2/+ mice have increased reticular fiber accumulation in skin compared WT littermates. Reticular fiber staining was performed on mice of the indicated ages (2-23 weeks) and strains. Stars mark the location of the epidermis. COL3A1 fibers (brown staining) are much thicker and more abundant at each stage of life in Tsk2/+ than in WT. Fibers were found to be especially pronounced in the panniculus carnosus region of the tissue; increased staining of COL3A1 in the dermis was also noted. All images were taken at 200X magnification.

Table 1. Nucleotide changes between Tsk2/+ mice and 101/H or B6 mice

Nucleotide position on chr 1 (mm9)	Genotype of Tsk2/+	Genotype of B6	Genotype of 101/H	Present in Other Strains?	Potential candidate for *Tsk2*?	Protein or mRNA containing substitution
SNP found by RNA-Seq						
44,675,490	A	T	T	No	No, outside interval	GULP1 Intron
44,833,682	C	T	T	No	YES	GULP1 Intron
45,378,353	A	T	T	No	YES	COL3A1 Exon (C33S)
45,432,389	C	G	nd	Yes	No	COL5A2 3'UTR
45,441,243	C	A	C	No	No, in 101/H	COL5A2 Intron
45,860,529	G	A	G	Yes	No	WDR75 Intron
45,874,790	T	C	T	Yes	No	WDR75 Intron
45,875,728	C	T	C	Yes	No	WDR75 Exon
45,880,257	CG	AC	CG	No	No, in 101/H	WDR75 Exon
46,872,610	T	G	nd	Yes	No	SLC39A10 Intron
46,874,711	C	T	C	Yes	No	SLC39A10 Intron
46,939,340	T	C	T	Yes	No	BC040767 Intron
46,939,624	A	G	nd	Yes	No	BC040767 Intron
SNP found by 454 Sequencing						
44,833,682	C	T	T	No	YES	GULP1 Intron
45,378,353	A	T	T	No	YES	COL3A1 Exon (C33S)
45,465,923	A	T	T	No	YES	Col5a2 Intron
46,124,856	A	G	A	Yes	No	Dnahc76 Intron
46,124,857	A	C	T	Yes	No	Dnahc76 Intron

All single-copy nucleotide changes found by RNA-Seq or 454 sequencing were checked for their presence in other non-fibrotic strains (http://phenome.jax.org/) or individually verified by a phototyping assay[34] and/or resequencing to confirm the single nucleotide change. SNP that were ruled out by one of these assays are considered not to be potential candidates for *Tsk2*. When known, genotypes shown for 101/H are from RNA-Seq, 454 sequencing or phototyping. Nd = not determined.

Table 2. Progeny born from *Col3a1*-deficient, *Col3a1*-sufficient, and Tsk2/+ mice.

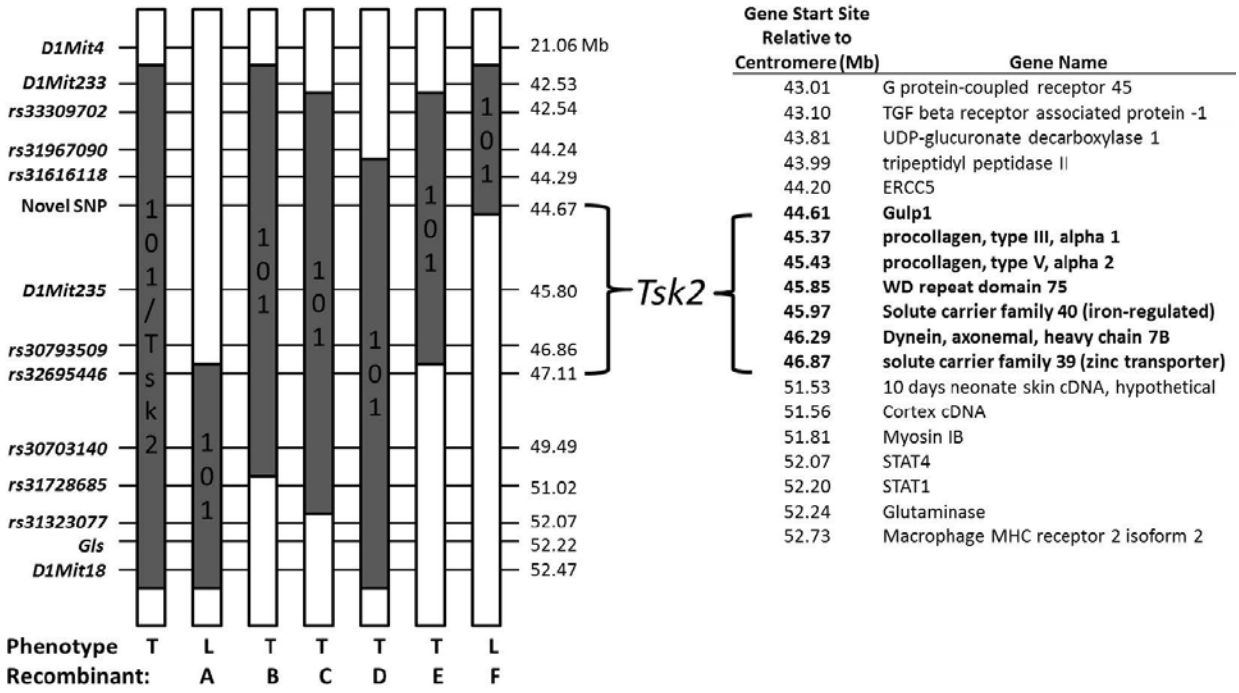
A.	Genotype and phenotype of progeny		
Parents	Tsk2/+ (Tight skin)	WT / WT (Normal skin)	Tsk2/Tsk2 (lethal)
Tsk2/+ x Tsk2/+	22	21	0
	$Col3a1^{+}$/ $Col3a1^{-}$ (Normal	$Col3a1^{+}$/ $Col3a1^{+}$ (Normal	$Col3a1^{-}$/ $Col3a1^{-}$
$Col3a1^{-/+}$ x	16	13	3

B.	Genotype and phenotype of progeny			
Parents	WT/$Col3a1^{+}$ (Normal skin)	Tsk2/$Col3a1^{+}$ (Tight skin)	WT/$Col3a1^{-}$ (Normal skin)	Tsk2/$Col3a1^{-}$
Tsk2/+ x $Col3a1^{-/+}$	12	10	11	0

Legend: All progeny were assessed for chromosome 1 markers (SNPS and microsatellites) that characterize the origin of the tested allele (*Tsk2* or *Col3a1*).

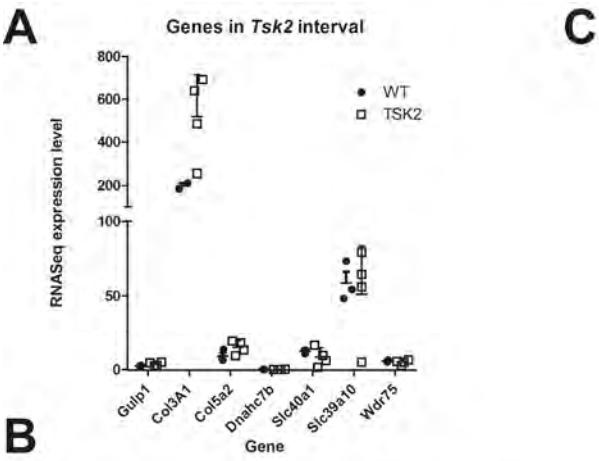
Table 2A (top) shows the number of mice born of each genotype and phenotype from Tsk2/+ x Tsk2/+ or $Col3a1^{-/+}$ x $Col3a1^{-/+}$ parents. Table 2B (bottom) shows the number of mice born of each genotype and phenotype from Tsk2/+ x $Col3a1^{-/+}$ parents; note there are no compound heterozygotes born from this mating.

Figure 1:



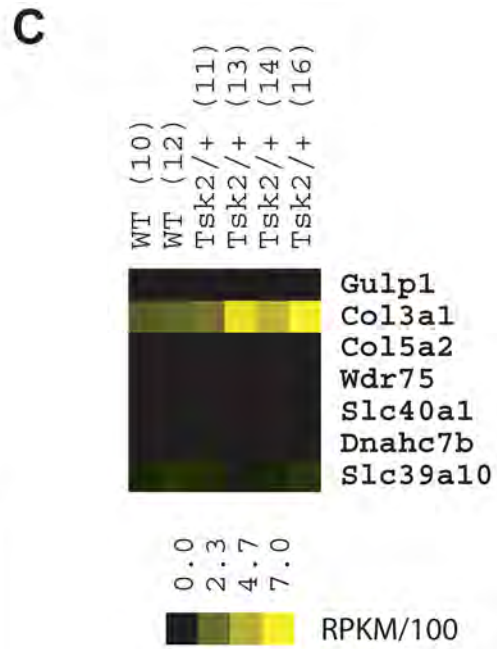
Gene Start Site Relative to Centromere (Mb)	Gene Name
43.01	G protein-coupled receptor 45
43.10	TGF beta receptor associated protein -1
43.81	UDP-glucuronate decarboxylase 1
43.99	tripeptidyl peptidase II
44.20	ERCC5
44.61	**Gulp1**
45.37	**procollagen, type III, alpha 1**
45.43	**procollagen, type V, alpha 2**
45.85	**WD repeat domain 75**
45.97	**Solute carrier family 40 (iron-regulated)**
46.29	**Dynein, axonemal, heavy chain 7B**
46.87	**solute carrier family 39 (zinc transporter)**
51.53	10 days neonate skin cDNA, hypothetical
51.56	Cortex cDNA
51.81	Myosin IB
52.07	STAT4
52.20	STAT1
52.24	Glutaminase
52.73	Macrophage MHC receptor 2 isoform 2

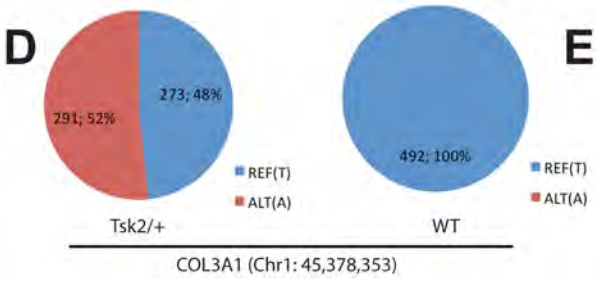
Figure 2:



B

Row Factor	Difference	t	P value	Summary
Gulp1	1.335	0.02951	P > 0.05	ns
Col3A1	320.5	6.249	P < 0.0001	****
Col5a2	5.919	0.1309	P > 0.05	ns
Dnahc7b	0.001221	2.700e-005	P > 0.05	ns
Slc40a1	-3.818	0.08441	P > 0.05	ns
Slc39a10	-7.491	0.1656	P > 0.05	ns
Wdr75	-0.7570	0.01674	P > 0.05	ns





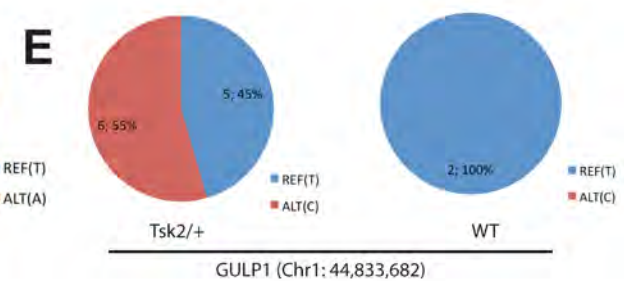


Figure 3:

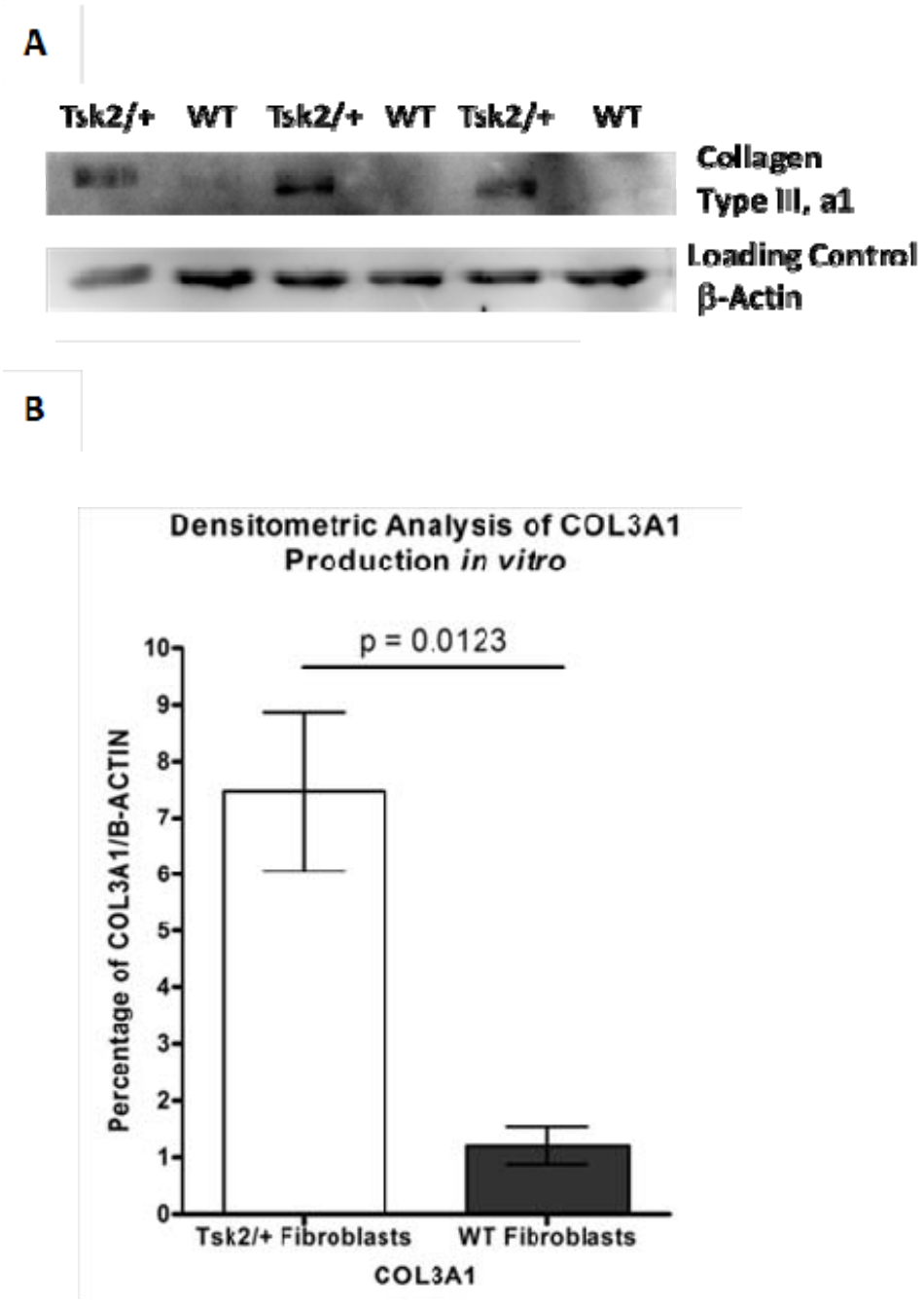
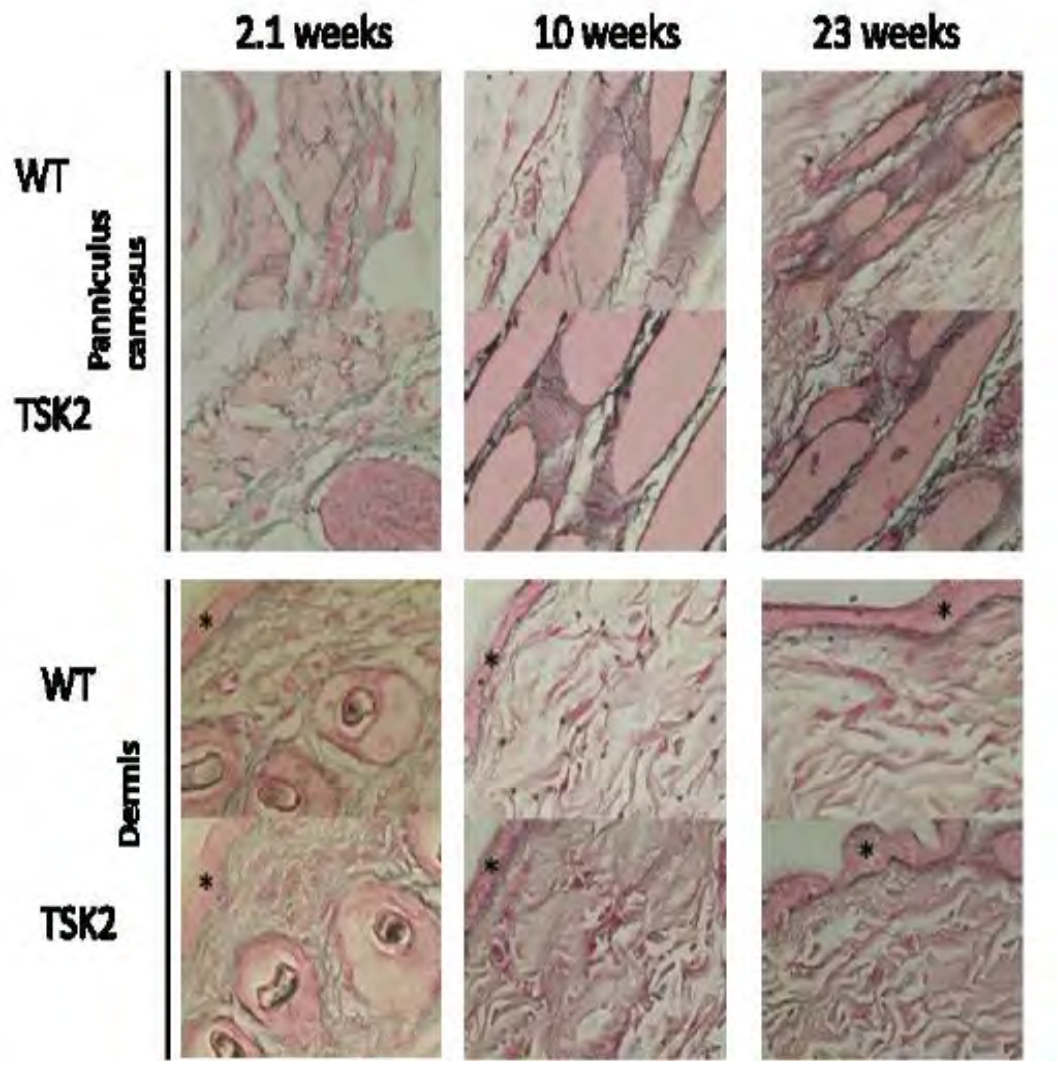

A
Tsk2/+ WT Tsk2/+ WT Tsk2/+ WT
Collagen Type III, a1
Loading Control β-Actin
B
Densitometric Analysis of COL3A1 Production in vitro
p = 0.0123
Percentage of COL3A1/B-ACTIN
0
1
2
3
4
5
6
7
8
9
10
Tsk2/+ Fibroblasts
WT Fibroblasts
COL3A1

Figure 4:


2.1 weeks
10 weeks
23 weeks
WT
Panniculus carnosus
TSK2
WT
Dermis
TSK2